DOCUMENT RESUME

ED 446 137 TM 031 899

AUTHOR	Lachat, Mary Ann
TITLE	What Policymakers and School Administrators Need To Know about Assessment Reform for English Language Learners.
INSTITUTION	Northeast and Islands Regional Educational Lab. at Brown Univ., Providence, RI.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	1999-00-00
NOTE	121p.
CONTRACT	RJ96006401
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Educational Assessment; Elementary Secondary Education; *Limited English Speaking; Minority Groups; *Performance Based Assessment; Second Language Learning; *Student Evaluation; *Test Use
IDENTIFIERS	Reform Efforts

ABSTRACT

This publication presents a comprehensive overview of one of the most profound shifts in educational policy and practice that has occurred during the 20th century--the transition from a testing culture to an assessment culture--and discusses its implications for English language learners (ELLs). The document brings together a wide range of research literature in a question and answer format. Chapter 1 discusses why assessment is viewed as a powerful tool for education reform, what the shift from testing to assessment means, the choices educators must make, and the implications of using a standards model for large-scale assessment programs. Chapter 2 discusses the characteristics of ELLS, how language and culture affect how ELLs learn, and the hopes and cautions of assessment reform for ELLs. Key issues are presented in chapter 3, including general and technical factors that influence equity in assessment of ELLS. Each chapter contains references. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

TM

1

ED 446 137

TM031899

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

2

BEST COPY AVAILABLE

Northeast and Islands Regional Educational Laboratory At Brown University (LAB)

The LAB, a program of The Education Alliance at Brown University, is one of ten federally supported educational laboratories in the nation. Our goals are to improve teaching and learning, advance school improvement, build capacity for reform, and develop strategic alliances with key members of the region's education and policy making community.

The LAB develops educational products and services for school administrators, policymakers, teachers, and parents in New England, New York, Puerto Rico, and the Virgin Islands. Central to our efforts is a commitment to equity and excellence.

Information about LAB programs and services is available by contacting:

LAB at Brown University
Education Alliance
222 Richmond Street, Suite 300
Providence, RI 02903-4226
Phone: 800-521-9550
Email: lab@brown.edu
Web: http://www.lab.brown.edu
Fax: 401-421-7650

The Center for Resource Management (CRM), based in South Hampton, New Hampshire, is a LAB partner organization.

About the author:

Mary Ann Lachat is president of CRM and program leader of the LAB's Standards, Assessment and Instruction Initiative.

Copyright © 1999, Brown University. All rights reserved.

This publication is based on work supported by the Office of Educational Research and Improvement (OERI), U.S. Department of Education, under Contract Number RJ96006401. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of OERI, the U.S. Department of Education, or any other agency of the U.S. Government.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

Contents

Preface i

Introduction iii

Chapter 1: A New Vision of Assessment

Why Is Assessment Viewed as a Powerful Tool of Education Reform? 7

What are the Differences between a Testing and an Assessment Culture? 14

What Choices Can Policymakers and School Administrators Make About the Use of Assessment Results? 18

How Will a Standards Model Affect Large-Scale State Assessment Programs? 21

Implications for Policy and Practice 30

Chapter 2: Assessment Reform and English Language Learners

Who Are the English Language Learners in America's Schools? 41

How Do Language and Culture Affect How English Language Learners Learn? 46

How Have Assessment Policies Affected the Education of English Language Learners? 51

Will Assessment Reform Help English Language Learners? 60

BEST COPY AVAILABLE

Chapter 3: Inclusive and Equitable Assessment for English Language Learners

What Factors Must be Considered in Order to Assess English Language Learners Equitably? 77

Will Performance Assessment Benefit English Language Learners? 81

What Policies and Practices Should be Followed When Including English Language Learners in Statewide Assessment Programs? 90

What Policies and Practices Should School Administrators and Teachers Follow When Assessing the Academic Performance of English Language Learners? 98

Implications for Policy and Practice 104

BEST COPY AVAILABLE

Preface

What Policymakers and School Administrators Need to Know About Assessment Reform for English Language Learners was produced to promote greater understanding of the significant issues that must be addressed to ensure inclusive and equitable assessment for linguistically and culturally diverse student populations. Its purpose is to translate the most important findings from the research literature into practical terminology, and to summarize the implications for policy and practice in ways that will be useful to state and local policymakers, superintendents, principals, school district personnel responsible for assessment, and bilingual, ESL, and Title I program directors.

Across the nation, America's classrooms are becoming increasingly diverse, and students whose first language is not English are the fastest-growing school population. Currently referred to as "English Language Learners" (ELLs), these children come from highly diverse backgrounds, and they face considerable challenges as they concurrently work toward English proficiency and respond to the academic demands of school. Assessment policies exert considerable control over the education of ELLs, from identification and classification through placement and ongoing monitoring of progress, shaping teacher beliefs about their abilities and the nature and quality of instruction offered to them. As noted in this publication, however, assessment practices in American schools were neither created nor designed to be responsive to the range of diversity represented in today's ELL student population, and in many ways have compounded inequities in their access to a high quality education.

While some educators feel that ELLs are over-tested, it is equally true that in many cases they have been under-assessed because much of what they know and can do has not been captured through traditional testing practices. Neither our national assessment programs nor most statewide assessment programs provide adequate data on the academic progress of ELLs.

This publication presents a comprehensive overview of one of the most profound shifts in educational policy and practice that has occurred during this century—the transition from a testing culture to an assessment culture—and discusses its implications for ELLs. Current reforms in assessment policies and practices have been viewed with some hope as important steps toward improving the quality of learning for all children, including ELLs. There are equal concerns, however, that development efforts have not sufficiently addressed the linguistic and cultural factors that impact on validity and fairness in assessment, nor issues of equity and access to the quality of instruction necessary to develop high level proficiencies.

The publication brings together a wide range of research literature in a question and answer format. Chapter 1 discusses why assessment is viewed as such a powerful tool of education reform; what it means to shift from a testing culture to an assessment culture; the choices that policymakers and school administrators must make about the purposes and uses of assessment; and the implications of using a standards model for large-scale state assessment programs. In Chapter 2, the following topics are discussed: the characteristics of ELLs in America's schools; how language and culture impact on how ELLs learn; how assessment policies have affected access to educational opportunity for ELLs; and the hopes and cautions of assessment reform for ELLs. Key issues as well as new visions of inclusive and equitable assessment policies and practices for ELLs are presented in Chapter 3, including general and technical factors that influence equity in assessment for ELLs; the advantages and cautions of performance-based assessment for ELLs; principles that should guide

the development of large-scale state assessments for ELLs; and what school administrators and teachers can do to ensure that school and classroom assessments of ELLs are appropriate.

Assessment policies must be consistent with our hopes for children and our vision of achieving both excellence and equity in the nation's schools. Developing and implementing sounder policies and practices will require policymakers and educational leaders to make new choices about the purposes and uses of assessment, challenge long-held beliefs about the capacity of diverse student populations to learn at high levels, and acquire greater awareness of how cultural and linguistic factors impact on learning. The findings of the researchers whose work is reflected in this publication provide important perspectives that can support and enhance efforts at state and local levels to ensure that assessment reform leads to positive results for all children.

Mary Ann Lachat, President
Center for Resource Management, Inc.*
2 Highland Road
South Hampton, NH

Program Leader for Standards,
Assessment and Instruction Initiative
Northeast and Islands Regional Educational
Laboratory at Brown University

*The Center for Resource Management, Inc. is a partner organization of the Northeast and Islands Regional Educational Laboratory at Brown University.

BEST COPY AVAILABLE

Introduction

With current educational reform initiatives calling for all students to attain high academic standards, national professional associations, states, and districts are moving swiftly to develop content standards specifying what American students should know and be able to do. Achievement of these content standards is to be measured through assessments, including performance assessments, based on the standards.

Yet data show that sizeable numbers of English language learners (ELLs) have routinely been exempted from state assessments. In 1994, my colleagues and I surveyed the 50 states and the District of Columbia to document policies concerning the participation of English language learners in statewide assessment programs. Of the 48 states responding, 44 reported allowing exemptions for ELLs. Exemptions are most often given pro forma on the basis of the student's English language proficiency or time spent in the U.S. ELLs were often allowed to be exempted for one to three years after arriving in the U.S. or in the school district (Rivera et. al., 1995). However, other reasons, such as teacher recommendation or participation in an ESL program, were sometimes given. Though meant to remedy the linguistic disadvantage ELLs face when taking English language content tests, the policy of exempting them creates a kind of systemic ignorance about their educational progress. The policy leaves

the school, district, or state unable to account for the learning of these students. In a reform climate where all students are expected to achieve to high standards, the inability of schools to be accountable for the success of ELLs is a constant reminder of the complexity of responding to the diverse educational needs of all U.S. students, including those learning English.

Given the complexity of involving ELLs in large-scale assessment programs, including state assessments, it is not surprising that states, districts, and schools feel challenged. For example, including ELLs in assessment programs, as research and experience shows, does not necessarily guarantee that meaningful information is collected on their progress. Often, ELLs are unable to demonstrate their knowledge and skills in content areas because of a lack of English proficiency. In such cases, assessments become tests of English proficiency as well as of the intended subject, with test scores yielding little more than an imperfect measure of the student's English language proficiency. The challenge for educators, therefore, is to create equitable systems which balance high quality and fair assessment strategies with the learning needs of English language learners.

Some progress, though, has been made in addressing the issue. Since 1994 many states have voluntarily moved to include ELLs in assessment programs. The U.S. Department of Education has also begun to develop policies that support the participation of ELLs in national assessments, such as the National Assessment of Educational Progress (Olson & Goldstein, 1997). These policies include the use of test modifications. However, the development of equitable assessment policies is a complex matter. While nearly all states now permit accommodations for ELLs when taking assessments, few accommodations are permitted in state-required high school graduation tests (Rivera & Vincent, in press). Also, while some states generally use a variety of accommodations, specific accom-

BEST COPY AVAILABLE

modation practices have raised concerns about the degree to which assessment policies contribute to educational equity.

The use of accommodations is not the only approach states and school districts are using. Many have developed or begun developing alternative assessments for ELLs. Arguably less dependent on English language skills, these assessments often allow varied means of letting students demonstrating whether a given standard has been attained. However, while these alternative assessments provide hope for an appropriate assessment tailored to the learner, their reliability and validity remain a source of concern.

Does putting the assessment in the student's native language balance the equity issue? Stansfield (1996) reported on the use of translations and adaptations in the context of state assessments. An adaption differs from direct translation in that it involves modifying test content in the process of translating the test. If carefully done, translations and adaptations can provide a more appropriate measure for some ELLs. However, the use of translated measures assumes that ELLs possess a considerable degree of literacy in their native language, which is not often the case. Translations and adaptations are especially appropriate if the student has been taught through his or her native language. In such cases, the student has had the opportunity to learn the academic language associated with the subject being assessed. However, few districts offer students such content instruction in their native language.

To create equitable assessments for large numbers of ELLs, the test development process must be reconceived with these learners in mind. Once the test and any alternatives are developed, appropriate test administration policies must be established and materials developed to guide educators in these policies. In short, to develop assessments and assessment policies that are equitable, educators must search for new strategies that meaningfully incorporate ELLs into state assessment programs.

BEST COPY AVAILABLE

An important first step in appropriately including ELLs in state assessments is to understand what the literature says about systemwide assessment and English language learners. One of only a few available reviews of literature on the assessment of ELLs, Mary Ann Lachat's monograph takes an objective look at what is currently known about systemwide assessment and ELLs, and draws recommendations about policies and procedures that can lead to the more equitable assessment of all students.

Her work should be particularly useful to assessment personnel in state and local education agencies. It will also be useful to teachers and content specialists who work on the development of assessments by encouraging them to develop test tasks that are more appropriate to a diverse student population. Further, it may give classroom teachers and content specialists ideas that they can implement in the development of alternative measures for ELLs.

Charlene Rivera, Director
Center for Equity and Excellence in Education
The George Washington University
Washington, DC

References

Olson, J. & Goldstein, A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress*. Washington, DC: U.S. Department of Education Office of Educational Research and Improvement.

Rivera, C. (1995). How can we ensure equity in statewide assessment programs? Findings from a national survey of assessment directors on statewide assessment policies for LEP students. Paper presented at the annual meeting of the National Conference on Large Scale Assessment, June 18, 1995, Phoenix, AZ. Washington, DC: George Washington University Center for Equity and Excellence in Education and Evaluation Assistance Center East.

Rivera, C. & Vincent, C. (In press). *High school graduation testing: Policies and practices in the assessment of English language learners*. Educational Assessment.

Stansfield, C.W. (1996). Content assessment in the native language. ERIC Digest, Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.

Chapter 1

A New Vision of Assessment

To understand how assessment reform affects students from different cultural backgrounds, we must understand the fundamental changes that are occurring in teaching and learning, testing and measurement, and school accountability. Amidst this sea of changes, the purpose and uses of assessment have been redefined.

> As a nation we do not all agree on the purposes of schools. Do we believe that schools are supposed to sort students to find the brightest and the best, or do we believe that our democracy will be stronger if we foster the creativity and capacity of every individual?. . . In choosing between the standards model and the measurement model, we will have made an implicit statement about what we believe to be the purpose of schools... The influences of each assessment model on our ways of thinking about learners and about our tasks as educators cannot be ignored. (Taylor, 1994)

Why Is Assessment Viewed As a Powerful Tool of Education Reform?

Assessment is a cornerstone of education reform. Across the nation, policymakers and educational leaders are employing new forms of assessment to improve the quality of education and to ensure accountability for

student learning. Four major factors make assessment a crucial part of education reform:

- Assessment reform powerfully affects equity and educational opportunity.
- There is a nationwide mandate for higher learning standards in America's schools.
- Contemporary research has altered our understanding of teaching and learning.
- We now recognize how much testing influences teachers and teaching.

PROMOTING EQUITY AND EDUCATIONAL OPPORTUNITY

Assessment reform is closely tied to the growing recognition that traditional testing practices have fueled inequities in education by relegating many students to a low-level education that limits their learning opportunities and life choices. Testing has traditionally been used to sort and rank students according to their abilities (presumed to be inherent) and then to track them into "appropriate" educational programs. In particular, testing practices have limited minority and low-income students' access to educational opportunities. Urban schools, which educate high proportions of students from low-income and varied cultural and language backgrounds, have disproportionately felt the negative impacts of testing policies (Darling-Hammond, 1991).

The purpose of assessment is shifting from deciding which students will have access to a high-quality education to ensuring that everyone will have the opportunity to achieve at high levels (Darling-Hammond, 1994). Based on cognitive research that shows that every individual possesses a range of knowledge and competence rather than a fixed level of ability (Resnick, 1987), new forms of assessment reflect a belief that tests should not penalize students or fail to accommodate diversity. By offering better ways of assessing the abilities of students who have underperformed on

traditional tests, reformers hope that new assessments will be used as tools for learning and student development, rather than tools for selecting which students should get the best educational opportunities (Farr & Trumbull, 1997; García & Pearson, 1994; Resnick & Resnick, 1992; Rothman, 1994).

Educators recognize, however, that in order for assessment reform to have a positive effect on the learning and achievement of students from low-income and culturally diverse populations, fundamental changes must occur in the fiscal policies that control the resources available to schools (Darling-Hammond, 1994; Winfield, 1995). For schools to educate students from all cultural and linguistic backgrounds equitably, they must give these students access to challenging curricula, resources, high-quality instruction, and a safe and supportive school environment. If these conditions are not in place, students will not achieve at high levels. New and promising approaches to instruction and assessment will not improve student achievement unless policies and practices that directly address inequities in learning conditions in schools are also put into effect (Neill, 1995; Stevens, 1996).

RAISING STANDARDS OF LEARNING FOR A 21ST CENTURY WORLD

Establishing high standards of learning for all students in America is the centerpiece of a national agenda to improve schools. Based on widespread recognition that many skills needed to function in today's world are not being taught in schools, reform efforts are defining the education standards essential for all students. Touching upon every aspect of the educational system, the movement to establish these standards is challenging long-held assumptions about how education should be conducted in our schools (Lachat, 1994).

Founded on the belief that it is in the national interest to educate all children and youth to their full potential, the standards movement aims to improve the quality of learning and teaching in America's schools and to break the cycle of failure experienced by so many of the nation's children. When children are not held to high academic standards, the results can be low achievement and the tragedy of students leaving school without ever having been challenged to fulfill their potential (Secretary of Labor's Commission on Achieving Necessary Skills [SCANS], 1991). By publicly defining standards for all students, schools can set clear, high expectations and establish a standard of education that does not deprive children of the chance to study a challenging curriculum or to have access to good jobs or further education when they finish school (National Council on Education Standards and Testing, 1992; National Education Goals Panel, 1993; SCANS, 1992). Public standards may also prod educators to avoid tracking students who are not fully proficient in English into limiting groups and to determine how these students can develop the knowledge and skills necessary for success in today's society.

The call for high learning standards has been paralleled by the demand for new assessment systems to measure their attainment—assessment systems that measure the achievement of higher-order cognitive abilities. Standardized tests were not designed to measure complex skills and performance abilities. As a result, they too often drive instruction toward lower-order cognitive skills (Darling-Hammond, 1994, 1995; Wolf, Bixby, Glenn, & Gardner, 1991). Testing practices in America have traditionally not matched new visions of equity and excellence, and in many cases testing has contributed to the educational problems that plague many schools. This is particularly true in schools that serve low-income and culturally diverse student populations. By seeking new assessment systems that will allow the country's diverse student populations to

BEST COPY AVAILABLE

demonstrate their ability to engage in complex tasks, educators and policymakers hope to foster the development of higher-order thinking skills, build upon the strengths and needs of individual learners, and encourage students to perform real world tasks (Resnick & Resnick, 1992; Shepherd, 1989; Taylor, 1994). Thus, new forms of assessment have an important role to play in a reformed education system "in which broader, more challenging, and more authentic educational values are operationalized and promoted" (García & Pearson, 1994, p. 337).

ASSESSMENT AND NEW CONCEPTS OF LEARNING

Recent research on the learning process has called into question the behaviorist view that learning is a sequential mastery of small skills that leads to the ability to perform higher-level activities. Instead, a cognitive perspective views intelligence as developmental and multifaceted, seeing learning as rooted in thinking and occurring through "performances of thought" that are characterized by uneven shifts in understanding involving multiple dimensions of intelligence (Gardner, 1993; Resnick, 1987; Resnick & Klopfer, 1989).

From this new perspective, learning is a constructive process that occurs through active knowing and thinking rather than through passive absorption of information. Learners actively construct their understanding of the tasks and situations they encounter. Research also suggests that a person's intellectual ability is not fixed, but can be enhanced by the learning process itself (Nickerson, 1989; Resnick, 1987; Sternberg, 1985; Wolf et al., 1991). Furthermore, learners develop as thinkers not in isolation, but by organizing and reorganizing knowledge while they interact with others and negotiate shared understandings (Resnick & Klopfer, 1989). "Understanding becomes deeper or more complex with the opportunity to witness other minds at work" (Wolf et al., 1991, p. 50).

Contemporary cognitive research has challenged common understandings about how and what children need to learn and invited educators to rethink how curriculum, instruction, and assessment are connected. It has also challenged outmoded theories of learning that in the past led to assessments that measured sequential rote instruction but not critical thinking. Based on this research, tests that focus on a narrow range of skills are being replaced with developmental performance assessments that reveal how students think and perform when solving complex problems (Baker, 1990; Resnick & Klopfer, 1989). Today, rather than teaching and then assessing isolated skills, teachers are starting to use assessment to support learning. They use assessments as an integral part of the learning process, having students solve problems by applying knowledge to real situations and allowing varied ways for them to demonstrate what they know and can do. These new approaches to teaching and learning may benefit students from different cultural backgrounds.

THE INFLUENCE OF ASSESSMENT ON TEACHING

Assessment exerts a powerful influence on teaching. Almost every state has some form of state-mandated testing program, and the testing industry affects students and teachers in every classroom in the nation. When large-scale testing programs were instituted to hold public schools accountable during the 1970s, teaching methods designed to develop higher order cognitive skills declined (National Center for Education Statistics, 1982). Increasingly, as teaching methods reflecting the lower cognitive demands of standardized tests became common in the nation's schools, only "the brightest and the best" students were encouraged to develop higher level cognitive abilities. In retrospect, the resulting decline in the academic performance of American students should have been anticipated. Major studies have since documented the negative impact of standardized

BEST COPY AVAILABLE

testing on teaching and learning in high schools (Goodlad, 1984; Sizer, 1985), and results of the 1992 and 1996 National Assessments of Educational Progress (NAEP) in mathematics showed that the majority of American students lag behind world-class standards of learning. Recent studies have underscored the international disadvantage created by the rote learning emphasized in U.S. classrooms (U.S. Department of Education, National Center for Education Statistics, 1996).

The belief that "high stakes" test scores were the most reliable indicator of both student achievement and educational quality has shaped educators' views about what should be taught in schools for decades. When an educational assessment provokes public scrutiny of test scores, educators feel pressured to improve how their students perform on that assessment by adapting instruction to mimic the demands of the test. As a result, tests exert an inordinate amount of influence on school curricula (Darling-Hammond, 1994; García & Pearson, 1994; Shepard, 1989; Smith, 1991). Simply stated, "You get what you assess" and "You do not get what you don't assess. . . . What does not appear on tests tends to disappear from classrooms in time" (Resnick & Resnick, 1992, p. 59). The practice of "teaching to the test" has been most pervasive in classrooms with a high percentage of students with low test scores, resulting in an over-emphasis on basic skills with the very students who would most benefit from a challenging curriculum.

Since testing influences teaching and learning so powerfully, education reform leaders advocate that new assessments be designed to have positive effects on classroom practice. "Assessments must be designed so that when teachers do the natural thing—that is, prepare their students to perform well—they will exercise the kinds of abilities and develop the kinds of skills and knowledge that are the real goals of educational reform" (Resnick & Resnick, 1992, p. 59).

What Are the Differences Between a Testing and an Assessment Culture?

Shifting from a testing to an assessment culture involves changing assumptions about the nature of intelligence and about how people learn. Because testing and assessment cultures have radically different belief systems and goals, helping educators and the public understand the implications of this change in point of view is an important part of education reform. A summary of the differences between the emerging assessment culture and the testing culture that dominated American education for half a century is shown in Figure 1.

The Impact of a Testing Culture on American Education

Based on a measurement model, the testing culture that has dominated American education for almost a hundred years assumed that intelligence and learning capacity were fixed traits that could be predicted. Because of this assumption, educators believed that students had an inherent level of intelligence which governed what they were able to learn. Therefore, the aim of testing was to sort and rank students for purposes of comparison and placement. Under the measurement model, the function of tests was to assess general knowledge across a broad range of achievement, to rank students based on their performance on the tests, and to compare students, schools, and districts on numeric achievement scales (Taylor, 1994).

The effects of ranking on American education have been wide reaching. Founded on early-twentieth-century theories that treat intelligence as a unitary, fixed trait, America's testing culture encouraged the belief that individuals could be ranked according to mental capacities. Because scores representing children's abilities were positioned relative to one another on a normal curve rather than determined by comparing performance to

FIGURE 1

School Policies and Practices That Support and Enhance Standards-Based Instruction and Assessment in Culturally and Linguistically Diverse Schools

School Policies:

- establish clear standards for what all students should know and be able to do;
- ensure that the curriculum offered to all students is based on the same standards for what students should know and be able to do;
- emphasize high expectations for all students — all students are provided with opportunities to achieve at high levels;
- ensure that all students are provided with equitable and adequate learning resources and high quality instruction;
- ensure sufficient time and resources for ongoing professional development to develop the teacher' capacities necessary for preparing diverse students to achieve at high levels
- reflect an understanding of the different purposes of assessment and the measures that are appropriate for these varying purposes;
- ensure that assessments are used for the primary purpose of improving student learning;
- emphasize equity and fairness in assessment for all students — assessments are not used to track or place students in narrow and limiting curricular programs or to inhibit educational opportunities.

Instruction and Assessment Practices:

- are based on desired learning results that are clearly understood by students, teachers, and parents;
- ensure that all students have adequate opportunities to develop higher order proficiencies;
- emphasize the integration of instruction and assessment;
- emphasize instruction that focuses on central concepts in depth rather than coverage of extensive information;
- emphasize an ongoing focus on student learning results for all students — every student has the opportunity to demonstrate achievement of learning standards during the school year;
- include provisions for identifying factors that might affect the performance of certain students or groups of students, and how these factors can be accommodated;
- draw upon the home and community experiences of culturally diverse students in developing authentic learning tasks;
- ensure that assessment results accurately reflect each student's actual knowledge, understanding, and achievement — assessments are designed to minimize the impact of biases on student performance;
- include procedures for determining the appropriateness of assessments for culturally and linguistically diverse students;
- include multiple measures and a variety of modes that allow culturally and linguistically diverse students to demonstrate what they know and can do;
- incorporate modifications that can be used in accommodating culturally diverse learners with varied levels of English proficiency;
- ensure that the development of classroom assessments includes a focus on how diverse students will be included in these assessments; and ensure that scoring rubrics are free of cultural bias and do not penalize students with varying levels of English proficiency.

24 BEST COPY AVAILABLE

criteria of achievement, "The result has been an enduring confusion between rank and accomplishment" (Wolf et al., 1991, p. 37). Because the normal curve assumes there will be only a few high performers, a large concentration of students in the middle, and a few poor performers, the use of the normal curve as the dominant profile for showing student achievement led to a widespread acceptance among policymakers, administrators, and teachers that a significant percentage of students would fail. Therefore, the belief system on which America's testing culture was based provided the public school system with a scientific rationale for tracking and blocked schools from confronting their responsibility for ensuring that all students learn and succeed academically. Testing helped schools to group students in classes according to their levels of ability and to design educational materials addressed to these different levels (Oakes & Lipton, 1990b; Wolf et al., 1991). High proportions of students who fell into the failing category under the traditional testing culture were poor students and students learning English as a second language. Because of their low test scores, these students were then placed in low-level classes.

A testing culture does not emphasize complex and rich ways of demonstrating learning. Focusing on a narrow range of cognitive abilities in order to magnify differences among students, a testing culture values accuracy, speed, and easily quantifiable skills. Because a test score takes on meaning only when compared to the scores of others, test items, which can range from very easy to very difficult, are selected for inclusion in a test based on how well they discriminate between high and low scores. An effective test of this type allows only a few examinees to score high so that scores can be easily differentiated and ranked (Farr & Trumbull, 1997; Taylor, 1994). Great value is placed on whether a testing instrument can predict ability and intelligence.

THE EMERGING ASSESSMENT CULTURE

The emerging assessment culture uses assessment as a tool to help schools and teachers learn about students rather than to classify, sort, or sanction them. Underlying this new approach to assessment is the belief that intelligence is not a fixed trait; instead, learning potential is considered to be developmental and a function of experience. Wolf et al describe an assessment culture as "defining and documenting what it is to use a mind well" (Wolf et al., 1991, p. 32), with an emphasis on informing teaching and learning rather than measuring and ranking students.

An assessment culture recognizes that intelligence is multifaceted—that people's multiple intelligences have varying degrees of strength and are at various stages of development. Therefore, intelligence cannot be accurately ranked according to a single dimension (Gardner, 1993). Assessment is treated as an "episode of learning" rather than something outside the learning process, and learning is understood to occur through both social interactions and individual reasoning (Neill, 1995). Since in this new model assessment becomes central to the instructional process, it is viewed as developmental, and student growth can be plotted in complex and rich ways (Resnick & Resnick, 1992). By shifting from a "measurement model" to a "standards model," assessment begins to focus on how student performance develops relative to standards of excellence, not on how each student ranks against other students.

The standards model is based on several important assumptions:

- Educators can agree upon standards of performance that will serve as learning targets.
- Most students can internalize and achieve the standards.
- Though student performances may differ, they will reflect the common standards.
- The standards defined by educators will allow for fair and consistent judging of diverse student performances.

(Taylor, 1994)

Because it is based on a standards model, the assessment culture emphasizes what students can do (student performances) not just what they know (content). Therefore, educators must not only define the content domain for their disciplines, but must also describe the complex performances and processes that are "authentic" to that discipline. Thus, the standards-based assessment culture has a new emphasis—on the collection of student work samples over time, on student performances that involve collaboration with others, and on assessing student work on complex problems that take an extended period of time to complete (Taylor, 1994; Wiggins, 1989, 1993).

It is hoped that the shift from a testing culture to an assessment culture will have a positive impact on the education of students who are learning English as a second language. Through the development of assessments based on clear standards of performance, educators may engage in a more open discussion about educational expectations for these students, about the quality of education offered to them, and about cultural bias and other factors that affect their performance.

What Choices Can Policymakers and School Administrators Make About the Use of Assessment Results?

At the center of the education reform debate lie questions about the choices policymakers and school leaders will make about the use of assessment results. Will they be used to determine student placements, reinforce differentiated curriculum tracking, and allocate rewards and sanctions to schools? Or will assessment results primarily be used to enhance teaching and learning and to increase educational opportunity for students who have traditionally been served poorly by public education? Today, be-

cause of financial and political demands, legislators and educators are demanding that the same assessment system serve incompatible purposes. These conflicting purposes have produced a tension at the heart of assessment reform.

> A perennial problem of testing programs is that policymakers and others wish to use a single instrument for a multitude of purposes–for example, to foster good teaching and learning, to make high-stakes decisions about individuals, to hold schools and districts accountable, to facilitate a voucher system, and to monitor national progress toward realizing federal, state, and local educational goals. Long experience with issues of test design, scoring, reporting, and the need for a supporting infrastructure teaches that these different purposes require different procedures and techniques. (Madaus, 1994, p. 88)

Policymakers and educational leaders must understand the implications of their choices, for new forms of assessment "will not be powerful or useful tools unless those who use them have a fundamental understanding of and belief in the views of learning and knowing to which they are conceptually linked" (Farr & Trumbull, 1997, p. 26). When policy decisions are made without clearly evaluating the intended purpose and use of assessment, unintended consequences that are destructive to children result; often, these consequences are particularly harmful to poor and language-minority children.

> At every level of analysis, assessment is a political act. Assessments tell people how they should value themselves and others. They open doors for some and close them for others. . . . The political dilemma is a problem for all students, but it is particularly acute for students from diverse cultural, linguistic, and economic backgrounds whose cultures, languages, and identities have been at best ignored and at worst betrayed in the assessment process. (García & Pearson, 1994)

28 BEST COPY AVAILABLE

For example, the testing provisions of Chapter I, which were legitimately established to determine accountability, unfortunately led to the use of low-level, multiple-choice tests for programmatic decisions and, as a result, to low-level instruction for poor children (Commission on Chapter I, 1992; Linn, Graue, & Sanders, 1990). Thus, because they were misused, basic skills tests were most damaging to the students they were intended to help. Similarly, minimum competency tests mandated by policymakers to ensure that all students had achieved basic math and literacy skills "corrupted" instruction by encouraging an emphasis on low-level skills and test preparation. Furthermore, because low-income and language-minority students disproportionately failed minimum competency tests, they were subsequently subjected to more intensive test preparation geared toward low-level skills. As a result, they were denied the opportunity to develop the capacities they would need to succeed in the future. In short, the quality of education made available to many students has been undermined by the testing policies and practices used to monitor and define their learning (Darling-Hammond, 1991, 1994; Haney, Madaus, & Lyons 1993; Madaus, 1991; National Commission on Testing and Public Policy, 1990; Oakes, 1985, 1986).

Standardized tests have long been used by schools to track students into different instructional programs. For most of the twentieth century, the IQ testing methods developed by Binet have been widely used to label and, frequently, to misclassify students. In many cases, African-American and Hispanic students have been disproportionately placed in dead-end classes (Gould, 1981; Madaus, 1994). Tests therefore serve as policy mechanisms that define educational opportunity and determine how students must demonstrate their competence (Darling-Hammond, 1995; Oakes, 1990).

> The role of testing in reinforcing and extending social inequalities in educational opportunities has by now been extensively researched and widely acknowledged. Use of tests for placements and promotions ultimately reduces the amount of learning achieved by students placed in lower tracks or held back in grade. Minority students are disproportionately subject to both of these outcomes of testing. (Darling-Hammond, 1995)

Today, tracking policies based on assessment systems continue to isolate many students from resources, the best teachers, and the best instructional practices. However, due to a growing intolerance of test policies that limit students' access to learning, the concept of "consequential validity" has emerged. Consequential validity stresses that an assessment's use is what matters—that is, whether the use of assessment results produces positive consequences for students and for the teaching and learning process (Farr & Trumbull, 1997; Shepard, 1993). It draws attention to the inequities produced when test results are used to limit educational opportunities. Consequential validity emphasizes that the use of assessment results is as important as technical concerns about reliability and content validity and that tests must be evaluated in terms of their effects on the lives of students.

How Will a Standards Model Affect Large-Scale State Assessment Programs?

States are increasingly turning to a standards-based model in developing statewide assessment systems that will be used to measure the progress of all students. These emerging systems represent a new way of thinking about large-scale assessment. For the first time, student learning will be measured against publicly defined standards of achievement, and performance-based assessment methods will be used to measure student

BEST COPY AVAILABLE

proficiencies. Assessment policies, rather than emphasizing ranking, are focusing on improving student learning, and as a result are creating new concepts of accountability for schools. Finally, technical criteria of validity and reliability are being re-defined.

MAKING STANDARDS OF ACHIEVEMENT PUBLIC

A standards-based model of large-scale assessment encourages educators and stakeholders to participate in defining the standards of quality toward which all students should strive. When students are measured against publicly defined standards of achievement, rather than against national norms established by test companies, public discussion of the appropriateness of any given standard for various student populations is possible (Lachat, 1994). Some see this increase in public participation in the setting of education standards as one of the most hopeful aspects of large-scale assessment reform. If the conception, development, and interpretation of assessment become open processes, then hidden biases will become more visible and more of the public will have a clear sense of what counts in our schools (García & Pearson, 1994).

By engaging educators and stakeholders in setting standards and in producing "Curriculum Frameworks" that organize education standards under major subject areas, state education agencies across the country are powering a nationwide curriculum reform movement by putting the standards model into effect. "The wager is that American education can be galvanized by setting high standards and using new, more probing assessments to hold districts, teachers, and students accountable. . . . [T]his bet is based on the hope that we can overcome past history and turn standards and testing into productive tools to guide reform" (Wolf et al., 1992). Two types of standards provide the foundation for standards-based assessment systems:

Content Standards define what children should know and be able to do. They describe the knowledge, skills, and understandings students should have in order to attain proficiency in a subject area. They describe what teachers are supposed to teach and what students are expected to learn. Content standards can serve as starting points for curriculum improvement because they describe what is important for all students in the various subject areas.

Performance Standards identify the levels a student can achieve in the subject matter defined in the content standards. They set specific expectations for various levels of proficiency and define what students must demonstrate to be considered proficient in the subject matter defined in the content standard. Performance standards are defined in terms of various levels of performance (a rubric). For example, a commonly used rubric in standards-based assessment systems defines student performance according to four levels: Advanced, Proficient, Basic, and Novice.

Many in education hope that the use of content and performance standards, by helping state officials, local educators, parents, and others agree on what students should learn, will create a clearer vision of academic success for all students in America's schools.

USE OF PERFORMANCE-BASED ASSESSMENTS

When assessments are tied to standards, students must demonstrate what they know and can do through a range of "performances," and new emphasis is placed on student work that involves higher order thinking and complex problem-solving. Performance-based assessments offer a better way of measuring the attainment of high learning standards and accommodating diversity than traditional assessments. Wiggins (1989) proposed that performance assessments were a more appropriate and meaningful

way of assessing student learning, suggesting that student performances that were "authentic" to the concepts, knowledge, and skills of a discipline and based on real world problems could be identified for all subject areas. When he first recommended that these identified performances should form the foundation of new assessment programs, his writings provoked strong response from policymakers and educators. At the time, though many in education were dissatisfied with standardized achievement tests, they also saw traditional tests as the only way to ensure fair and reliable large-scale testing (Taylor, 1994). Today, however, performance-based assessment is gradually becoming accepted as a promising vehicle for improving statewide assessment programs. An analysis of 1995 state assessment systems showed that 17 states were using performance assessments (*Education Week*/Pew Charitable Trusts, Special Report on the Condition of Education in the 50 States, 1997).

Performance-based assessment has been described as having the following key features:

- It compares student achievement to agreed-upon levels of proficiency or excellence.
- It solicits higher order thinking processes.
- It emphasizes the importance of context through assessment tasks reflecting real-life problems that are meaningful to the learner.
- It invites students to solve problems or performance tasks of varying complexity, some of which involve multiple steps, several types of performance, and significant student time.
- It sometimes demands both group and individual performance in responding to a task.

(Baker, O'Neil, & Linn 1993; Valdez-Pierce & O'Malley, 1992)

Both the design and interpretation of performance assessments rely on the judgement of those scoring the tests. The assessor must apply clearly articulated performance criteria in making a professional judgment about

the level of proficiency demonstrated. Scoring looks beyond right or wrong answers; it also considers the thoughtfulness of the procedure used to carry out the task or solve the problem (Shavelson, Baxter, & Pine, 1992).

A NEW POLICY PERSPECTIVE

The trend toward basing large-scale assessments on performance standards challenges practices that have dominated public schools for more than a half century. New assessments are being designed to stimulate student growth rather than to determine whether students are ready to profit from a high-level education; these assessments exchange "assessment for ranking" for "assessment to improve student learning." By changing assessment content (to knowledge and skills that are based on standards) and form (to tasks that invite complex performances), performance-based assessments significantly alter how students demonstrate what they know and can do. Many educators and policymakers believe that using performance-based assessments in emerging state assessment programs will change approaches to learning that have been based on measurement-oriented, multiple-choice testing. However, to fulfill the promise of this new perspective on assessment, strong leadership in policy and education practice will be needed; it will not be easy for either educators or the public to exchange measurement-driven assumptions for a new set of assumptions about the role of assessment in learning.

When considering changes in assessment policy, policymakers and educational leaders should attempt to anticipate the results of new approaches, for past assessment policies have sometimes had unintended consequences. For example, considerable research substantiates that the high-stakes testing programs of the 1980s narrowed the range of instruction in schools and even the scope of content covered in the tests themselves (Darling-

Hammond, 1991; Jaeger, 1991; Madaus, 1991; McLaughlin, 1991; Shepard, 1991). The lessons learned from the unanticipated results of previous testing initiatives should inform decision-making about new assessment programs that seek to be both more conducive to student learning and more equitable for diverse student populations.

NEW CONCEPTS OF ACCOUNTABILITY

Because standards-based assessments are part of a push toward higher levels of learning, they contribute to new demands for schools to verify that all students, including students who are not fully proficient in English, are achieving at acceptable levels. However, when policymakers link higher standards of performance to school accountability, they provoke considerable discussion and debate. To many educators, the drive for accountability exemplifies the kind of top-down approach to educational change that undermines reflective practices in teaching and learning. Other educators see accountability as a necessary part of current efforts to reform schools. At the heart of the debate is the widespread recognition that even if external authorities establish higher standards and provide inducements, many schools will still lack the organizational capacity to get their students to achieve at high levels (Newmann, King, & Rigdon, 1997). Many schools lack the necessary resources to respond to the needs of their increasingly diverse student populations (Baker, 1992). Therefore, when schools are held accountable for ensuring that all students achieve high standards of learning, many complex issues are raised about the inequities that exist between schools.

Concerns about equity in school resources have been expressed in the literature on school reform, and several necessary components of organizational capacity have been identified:

- Teachers must have the knowledge, skills, and capabilities to provide high-quality instruction.
- Administrators must provide effective leadership.
- Financial resources and programmatic resources—including curriculum and assessment materials that support high levels of learning, laboratories, libraries, and computing facilities—must be available.
- Teachers and administrators must have access to high-quality professional development.
- The school environment must be safe and secure.
- Schools must have the organizational autonomy necessary for responding to the demands of the local context.

(Clair, Adger, Short, & Millen, 1998; Corcoran & Goertz, 1995; Darling-Hammond, 1993, 1995; Newmann et al., 1997; O'Day, Goertz, & Floden, 1995).

There are also widespread concerns about the inequitable effects of using penalties to enforce accountability when standards-based assessment systems are used. Darling-Hammond (1994) and others argue that rewards and penalties create powerful incentives for schools to restrict the participation of those students who might perform at lower than proficient levels. As a result, some accountability measures can undermine both the unique value of performance-based assessment systems and the larger attempt to provide all students in a diverse population with a high-quality education. "To protect the integrity of authentic assessments, we need to engage in thoughtful, ongoing conversations to determine what we gain and lose by making authentic assessments part of rigorous, high-stakes accountability" (Zessoules & Gardner, 1991, p. 70).

Wolf and her colleagues (1992) stress that the notion of accountability, while necessary for effective change to occur, should be envisioned in far richer and more complex ways than it is in typical state mandates.

While schools owe their constituencies honest accounts of what they have and have not achieved, narrow visions of accountability often result in assessment becoming driven too exclusively by concerns for measuring and reporting achievement data for outside audiences. Wolf et al. (1992) assert the importance of "internal accountability"—encouraging students, teachers, and families to reflect on what is worth knowing and ensuring that all students have the opportunity to develop essential knowledge. They see internal accountability as a more appropriate focus for the reporting and interpreting of assessment results than external mandates.

TECHNICAL CRITERIA FOR VALIDITY AND RELIABILITY

The increasing use of large-scale performance-based assessment has forced experts who deal with the technical aspects of assessment to rethink their methods for assuring quality. The litmus test for any measuring instrument has always been its degree of reliability (the degree to which the test yields the same results on repeated trials) and of validity (the degree to which a test measures what it is intended to measure and to which inferences made based on a test's results are appropriate and useful). In the past, the testing culture was prone to sacrifice validity to achieve reliability, in effect sacrificing the student's interests for the test maker's (Wiggins, 1993). However, new approaches to assessment have led some to question the role reliability has traditionally played in assessment. Because content validity, or the ability to understand what student performance reveals about learning, is of primary importance in performance assessment, Wolf et al. (1991) suggest that we "revise our notions that high-agreement reliability is a cardinal symptom of a useful and viable approach to scoring student performance" (p. 63).

Because performance assessments, by their nature, often require integrated knowledge and skills, they are far less standardized than traditional

BEST COPY AVAILABLE

tests and allow for more latitude in design, in student response, and in scorer interpretation. As a result, establishing reliability has been a major issue in large-scale performance assessments. There has been more success in establishing consistency in scoring among well-trained raters than in establishing consistency across tasks (Baker et al., 1993). For example, developers have had difficulty in establishing acceptable levels of comparability (reliability) across tasks intended to address the same skills (Farr & Trumbull, 1997). Some research shows that consistency of performance across tasks is influenced by the extent to which tasks both share comparable features and reflect the types of instruction students had received. Research also has shown that variations in task performance may be attributable to differences in students' prior knowledge and their experiences in performing similar tasks (Linn, Baker, & Dunbar 1991; Shavelson, et al., 1992).

Farr & Trumbull (1997) suggest that "the tension between validity and reliability that arises when standardization is reduced may resolve itself in the direction of validity, particularly if an integrated view of students' performances and learnings takes the place of a focus on individual samples of performance out of context" (p. 56). They cite the important work of Messick (1989) and Cronbach (1989) in broadening the definition of validity to include the social consequences of assessment and also cite the work of validity researchers who have sought to balance concerns about reliability and generalizability with consideration of additional criteria such as "authenticity" (Newmann, 1990) and "cognitive complexity" (Linn et al., 1991).

Researchers have given increasing attention to the validity criteria that should characterize the use of performance assessments in large-scale statewide assessment programs (Baker et al., 1993; Linn et al., 1991; Messick, 1989). Leading assessment specialists recommend that these assessments

exemplify current content standards for what students should know and be able to do in various subject areas, and also contain explicit standards for rating or judging performance. The assessments should require that complex cognition be demonstrated through knowledge representation and problem solving. The validity criteria recommended by assessment specialists also stress that performance assessments be fair to students of different backgrounds and meaningful to students and teachers, incorporating competencies that can be taught and learned. The emergence of large-scale performance assessments has thus highlighted the importance of establishing a connection between validity standards and the policy uses of assessment.

IMPLICATIONS FOR POLICY AND PRACTICE

Provide leadership that helps schools make the transition from a testing culture to an assessment culture.

The transition from a testing culture to an assessment culture represents the most profound shift in education policy and practice that has occurred during this century. Because testing cultures and assessment cultures are based on radically different belief systems, educators and the public need help in understanding what this transition means. Educators, stakeholders, and the public at large must develop entirely different assumptions about learning, the nature of intelligence, and the purposes of assessment (Wolf et al., 1991). The value system underlying a "ranking and comparing" model of assessment has had a powerful influence on the thinking of educators and the public at large. Shifting to a new paradigm will require changing widespread beliefs about children's innate abilities and capacities to learn (Madaus, 1994).

Give priority to developing assessments that send new signals about what children need to learn.

Because assessment shapes teacher beliefs about what should be taught, assessments must support what we want teachers to teach. If policymakers endorse high standards for student learning as the foundation for large-scale assessment of students, these standards will also drive instructional improvement. Standards-based assessment reform may thus motivate schools to design curricula around the key concepts, principles, understandings, and skills that all children should have the opportunity to learn (Resnick & Resnick, 1992; Wolf et al., 1992). Such reform may improve classroom instruction for previously underserved student populations (like English language learners) by encouraging teachers to revise their teaching of these students to emphasize complex thinking and problem-solving processes in regular classroom activities (García & Pearson, 1994).

Provide leadership that helps educators and the public understand and accept new assumptions about the role of assessments.

Policymakers and educational leaders will have to make clear choices about the purposes and uses of assessment and build support for the belief that all students are capable of learning and achieving at high levels. Legislators and educators will have to decide whether they want assessment systems that will be used to rank and compare students, schools, and districts; or whether they want assessment systems that will be used to guide and measure student progress toward desired standards of excellence. They will also need to decide whether they want assessment systems that select and serve the brightest and the best or that enhance the learning of all children (Farr & Trumbull, 1997; Madaus, 1994; Taylor, 1994).

Develop assessment policies that address the dual goals of achieving both excellence and equity in the nation's schools.

It is still not known how new assessments will affect those students in schools with the least supportive environments and from non-mainstream

cultural and linguistic backgrounds (García & Pearson, 1994). Madaus (1994) suggests that policymakers and educational leaders keep the following questions in mind to minimize the side effects associated with previous policy-driven testing programs: What are the unintended consequences that may result from the use of new assessments? How will new assessments affect groups traditionally disadvantaged by tests? How can we minimize possible negative effects on students from minority populations and from diverse cultural and linguistic backgrounds?

Use the principle of consequential validity as a benchmark for making decisions about the purpose and uses of assessment.

The construct of consequential validity emphasizes that assessment results are not valid either when students' abilities and potential have not been judged fairly because of the inappropriateness of the assessments for the population being assessed, or when the use of test results deprives certain students from having access to the best teachers and to high quality learning environments. When tests are used to make important educational decisions, test limitations and misuse become more damaging. Basing important decisions on flawed assessments has resulted in particularly negative consequences for students from non-mainstream cultural and linguistic backgrounds. Future policies must ensure that tests are used appropriately for all students (Estrin, 1993; Koelsch, Estrin, & Farr, 1995; García & Pearson, 1994; Shepard, 1993).

Help schools address the significant challenges of standards-based reform with highly diverse student populations.

Never before have our schools been asked to ensure that all students achieve publicly defined standards of learning. Never before have we asked schools to consider "higher-order" skills as core skills that all students, not just the most gifted, need to acquire. Never before have teachers faced such diverse and challenging student populations (Lachat, 1994). To change

to a standards-based model, schools will have to break free of assumptions about differential abilities in learners and address the differential conditions that affect learning. At the same time, schools will need to provide ongoing, in-depth professional development for teachers. Furthermore, schools will need to find new ways to respond to community needs for accountability while working with political and community leaders more intensively than ever before to address the impact of social and economic conditions on children (Farr & Trumbull, 1997; Taylor, 1994).

Make significant investments in professional development.

Preparing teachers to use recognized best practices when educating diverse student populations requires significant investments in professional development. Educational approaches that sort children into those who have access to high-level learning and those who will focus only on basic skills and simple tasks are no longer seen as valid (Farr & Trumbull, 1997). New approaches to educational practice emphasize how curriculum, instruction, and assessment relate to one another, and many educators now view teaching, learning, and assessment as inextricably linked. However, many teachers have little knowledge of the strategies that support new models of instruction, and many hold beliefs and priorities that are incompatible with the positive changes envisioned for the nation's schools (Oakes & Lipton, 1990; Resnick & Resnick, 1992). Because new teaching and learning models have not yet spread to most of the nation's schools, professional development is a key to advancing improvements in classroom instruction.

Actively involve groups at state and local levels in creating public understanding of large-scale assessments based on high standards of learning.

Schools have always been faced with the demands of many groups including their local school board, district administration, state and federal

agencies, parents, politicians, and the business community. In the past, the expectations of these groups have varied considerably (Newmann, et al., 1997). For large-scale standards-based assessment systems to serve the goal of high levels of learning for all students, various groups of education stakeholders will need to have a shared vision and provide coordinated support to education reform.

References

American Psychological Association, American Educational Research Association, National Council of Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Baker, E.L. (1990). Developing comprehensive assessments of higher order thinking. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics* (pp. 7-20). Washington, DC: American Association for the Advancement of Science.

Baker, E.L. (1992). Issues in policy, assessment, and equity. In *Focus on evaluation and measurement*, volumes 1 and 2. Proceedings of the National Research Symposium on Limited English Proficient Student Issues (2nd ed., Washington, DC, September 4-6, 1991).

Baker, E.L., O'Neil, H.R., Jr., & Linn, R.L. (1993). Policy and validity prospects for performance-based assessment. *American Psychological Association*, 48 (12), 1210-1218.

Commission on Chapter I. (1992). *High performance schools: No exceptions, no excuses*. Washington, DC: Author.

Corcoran, T., & Goertz, M. (1995). Instructional capacity and high performance schools. *Educational Researcher*, 24 (9), 27-31.

Clair, N. Adger, C., Short, D., & Millen, E.M. (1998). *Implementi ig standards with English language learners: Initial findings from four middle schools*. Providence, RI: Northeast and Islands Regional Educational Laboratory at Brown University.

Cronbach, L.J. (1989). Construct validation after thirty years. In R.L. Linn (Ed.), *Intelligence: Measurement theory and public policy*. Urbana, IL: University of Illinois Press.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappan*, *73*, 220-225.

Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan*, 74, 753-761.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64, 5-30.

Darling-Hammond, L. (1995). Inequality and access to knowledge. In J. Banks & C. Banks (Eds.), *Handbook of research on multicultural education* (pp. 465-483). New York, NY: Macmillan Publishing USA.

Farr, B., & Trumbull, E. (1997). *Assessment alternatives for diverse classrooms*. Norwood, MA: Christopher-Gordon.

García, G., & Pearson, P. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of Research in Education, 20* (pp. 337-391). Washington, DC: American Educational Research Association.

Gardner, H. (1993). *Frames of mind*. New York, NY: Basic Books.

Goodlad, J. (1984). *A place called school: Prospects for the future*. New York, NY: McGraw-Hill.

Gordon, E. (1992). *Implications of diversity in human characteristics for authentic assessment.* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Gould, S.J. (1981). *The mismeasure of man.* New York, NY: W.W. Norton.

Haney, W., Madaus, G.F., & Lyons, R. (1993). *The fractured marketplace for standardized testing.* Boston: Kluwer Academic Publishers.

Jaeger, R.M. (1991). Legislative perspectives on statewide testing: Goals, hopes, and desires. *Phi Delta Kappan*, 73 (3), 239-242.

Koelsch, N., Estrin, E., & Farr, B. (1995). *Guide to developing equitable performance assessments.* San Francisco: WestEd.

Lachat, M.A. (1994). *High standards for all students: Opportunities and challenges.* South Hampton, NH: Center for Resource Management, Inc.

Linn, R.L., Graue, B., & Sanders, N.M. (1990). Comparing state and district test results to national norms: Interpretations of scoring "above the national average." *Educational Measurement: Issues and Practices*, 9 (3), 5-14.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20 (8), 1521.

Madaus, G.F. (1991). The effects of important tests on students: Implications for a national examination or system of examinations. *Phi Delta Kappan*, 73 (3), 226-231.

Madaus, G.F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Educational Review*, 1 (64), 76-95.

McLaughlin, M.W. (1991). Test-based accountability as a reform strategy. *Phi Delta Kappan*, 73 (3), 248-251.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.

National Center for Education Statistics. (1982). *The conditions of education, 1982.* Washington, DC: U.S. Department of Education.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America.* Chestnut Hill, MA: Boston College.

National Council on Education Standards and Testing. (1992). *Raising standards for American education.* Washington, DC: Author.

National Education Goals Panel. (1993). Setting standards, becoming the best. Chapter 1, Vol. 1 of the *1993 Goals Report.* Washington, D.C.

Neill, M. (1995). Some prerequisites for the establishment of equitable, inclusive multicultural assessment systems. In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 115-157). Boston: Kluwer Academic Publishers.

Newmann, F.M. (1990). Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies*, 22, (1), 41-56.

BEST COPY AVAILABLE

Newmann, F.M., King, M.B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review, 1, (67)*, 41-74.

Nickerson, R.S. (1989). On improving thinking through instruction. In Cazden, C.B. (Ed.), *Review of Research in Education* (pp. 3-57). Washington, DC: American Educational Research Association.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Oakes, J. (1986). Tracking in secondary schools: A contextual perspective. *Educational Psychologist, 22*, 129-154.

Oakes, J. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science. Santa Monica, CA: RAND.

Oakes, J., & Lipton, M. (1990). Tracking and ability grouping: A structural barrier to access and achievement. In J. Goodlad & P. Keating (Eds.), *Access to knowledge: An agenda for our nation's schools* (pp. 187-204). New York: The College Entrance Examination Board.

O'Day, J., Goertz, M.E., & Floden, R.E. (1995). Building capacity for education reform. *In CPRE Policy Briefs: Reporting on issues and research in education policy*. New Brunswick, NJ: Rutgers University.

Resnick, L.B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

Resnick, L.B. & Klopfer, L.E. (Eds.). (1989). *Toward the thinking curriculum: Current cognitive research*. Washington, DC: Association for Supervision and Curriculum Development.

Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for education reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Rothman, R. (1994). *Assessment questions: Equity answers. Proceedings of the 1993 CRESST Conference*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Secretary of Labor's Commission on Achieving Necessary Skills (SCANS). (1991). *What work requires of schools: A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

Secretary of Labor's Commission on Achieving Necessary Skills (SCANS). (1992). *Skills and tasks for jobs*. Washington, DC: U.S. Department of Labor.

Sizer, T. (1985). *Horace's compromise: The dilemma of the American high school*. Boston: Houghton Mifflin.

Shavelson, R.J., Baxter, G.P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21*, 22-27.

Shepard, L.A. (1989). Why we need better tests. *Educational Leadership, 46*, 4-9.

Shepard, L.A. (1991). Will national tests improve student learning? *Phi Delta Kappan, 73* (3), 232-238.

BEST COPY AVAILABLE

Shepard, L.A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of Research in Education, 19*, (pp. 405-450). Washington, DC: American Educational Research Association.

Smith, M. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*, 8-11.

Sternberg, R. (1985). Beyond IQ. New York: Cambridge University Press.

Stevens, F.I. (1996). Closing the achievement gap: Opportunity to learn, standards, and assessment. In B. Williams (Ed.), *Closing the achievement gap: A vision for changing beliefs and practices* (pp. 77-95). Alexandria, VA: Association for Supervision and Curriculum Development.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal, 31*, 231-262.

U.S. Department of Education, National Center for Education Statistics. (1996). Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context. NCES 97-198, Washington, DC: Government Printing Office.

Valdez-Pierce, L., & O'Malley, J.M. (1992). *Performance and portfolio assessment for language minority students*. Washington, DC: National Clearinghouse on Bilingual Education.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46, 41-47.

Wiggins, G. (1993). *Assessing student performance*. San Francisco: Jossey-Bass.

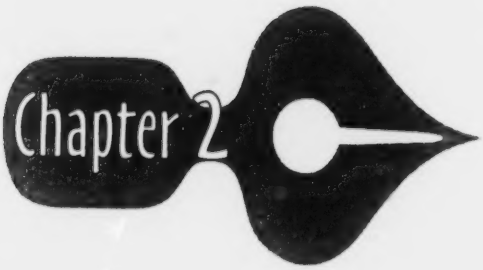
Winfield, L. (1995). Performance-based assessments: Contributor or detractor to equity? In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment*. Boston: Kluwer Academic Publishers.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of Research in Education* (pp. 31-74). Washington, DC: American Educational Research Association.

Wolf, D.P., LeMahieu, P.G., & Eresch, J.A. (1992). Good measure: Assessment as a tool for educational reform. *Educational Leadership*, 49, (May), 8-13.

Zessoules, R., & Gardner, H. (1991). Authentic assessment: Beyond the buzzword and into the classroom. In V. Perrone (Ed.), *Expanding student assessment* (pp. 47-71). Arlington, VA: Association for Supervision and Curriculum Development.

BEST COPY AVAILABLE


Chapter 2

Assessment Reform and English Language Learners

Those who make decisions about how to assess English language learners must take into account the characteristics of these students, the influence of language and culture on their learning, the factors in the daily life of schools that affect their ability to learn, and the ways that assessment policies control their access to a high-quality education. These issues provide the context for examining the potential influence of the current assessment reform movement on the education of English language learners.

> Assessment policy is not about whether to include, exclude, or exempt English language learners from assessments. Rather, the discussion must center around two questions: how best to assess English language learners, and how best to incorporate the data into accountability assessments of schools and school systems. (LaCelle-Peterson and Rivera, 1994)

Who Are the English Language Learners in America's Schools?

The term "English language learner" (ELL) is a recent designation for students whose first language is not English. This group includes students who are just beginning to learn English as well as those who have already

developed considerable proficiency. The term reflects a positive focus on what these students are accomplishing—mastering another language—and is preferred by some researchers to the term "limited English proficient" (LEP), the designation used in federal and state education legislation and most national and state data collection efforts (August & Hakuta, 1997; LaCelle-Peterson & Rivera, 1994).

The English language learner population is highly diverse, and any attempt to describe accurately the group as a whole, as with any diverse group of people, is bound to result in inaccurate generalizations. While this group of students share one important feature—the need to increase their proficiency in English—they differ in many other important respects. English language learners are a diverse cross-section of the public school student population. The primary language, cultural background, socio-economic status, family history, length of time in the United States, mobility, prior school experiences, or educational goals of any student in this group can distinguish him or her from any other English language learner. Often, common assumptions about students whose primary language backgrounds are in languages other than English are not accurate. For example, many think that the vast majority of English language learners are immigrants or recent arrivals to this country; however, according to a U.S. Department of Education report, 41% of English language learners were born in the United States.

> The umbrella of "English language learner" includes students from Native American communities that have been in what is now the United States from time immemorial; students from other long-established language minority communities, such as Franco-Americans in the Northeast, Latino and Chicano in the Southwest, and the Amish in the Midwest; and students from migrant and immigrant groups who represent the most recent arrivals in a virtually unbroken series of migrations that have brought linguistic diversity to North America. (LaCelle-Peterson & Rivera, 1994, p. 59)

BEST COPY AVAILABLE

It is difficult to accurately determine the number of English language learners in the nation's schools for a number of reasons. First, educators do not agree about what constitutes English language proficiency. Furthermore, effective assessment tools or practical methods for surveying the English skills of all students do not exist. Finally, different states—and even different districts within states—vary widely in the processes and standards they use to identify English language learners. For example, four recent surveys indicate that states and local districts use a variety of methods to determine which English language learners have limited English proficiency, to place students in language-related programs, and to monitor student progress in these programs (August & Lara, 1996; Cheung, Clements, & Mieu, 1994; Fleishman & Hopstock, 1993; Rivera, 1995). The variation among states means that a student could be considered to demonstrate limited English proficiency by one state but not by another (Rivera, Hafner, & LaCelle-Peterson, 1997). As a result of this inconsistency in categorizing student skills, state and national education policy analysts do not have reliable information about the numbers of students in need of language-support services (Clements, Lara, & Cheung, 1992).

An advisory committee to the Council of Chief State School Officers (CCSSO) developed the following definitions for describing the English proficiency of English language learners:

A *fully English proficient (FEP) student* is able to use English to ask questions, to understand teachers and reading materials, to test ideas, and to challenge what is being asked in the classroom. Four language skills contribute to proficiency:

- ***Reading***–the ability to comprehend and interpret text at the age- and grade-appropriate level.

- ***Listening***–the ability to understand the language of the teacher, to comprehend and extract information, and to follow the instructional discourse through which teachers provide information.
- ***Writing***–the ability to produce written text whose content and format fulfill classroom assignments at the age- and grade-appropriate level.
- ***Speaking***–the ability to use oral language appropriately and effectively in learning activities within the classroom (such as peer tutoring, collaborative learning activities, and question/answer sessions) and in social interactions within the school.

A *limited English proficient (LEP) student* has a language background other than English, and his or her proficiency in English is such that the probability of the student's academic success in an English-only classroom is below that of an academically successful peer with an English-language background.

The English language learner population is the fastest-growing subgroup of the school-age population today. Their growing numbers reflect demographic trends that have been occurring over the past twenty years. Several studies confirm the steady increase of English language learners in public schools.

- According to one study, during the 1991–92 school year 2.3 million K–12 students categorized as having limited proficiency in English were enrolled in public school districts. This figure indicates an increase of 1 million students over a 10-year period (Fleischman & Hopstock, 1993).
- In the 1993-94 school year, three million students categorized as having limited proficiency in English were identified out of a total of 45.4 million students enrolled in the U.S. public schools in the fifty states and the District of Columbia (Donly, Henderson, & Strang 1995). However, these statistics are considered to be conservative, given that, according to the 1990 census, over 6.3 million people between five and seventeen years of age (13.9% of all school-aged people in the nation) reported speaking a language other than English in their home (Waggoner, 1992).

BEST COPY AVAILABLE

- In another longitudinal analysis drawing upon multiple data sources including the U.S. Department of Education, the National Center for Education Statistics, and state sources, Olsen (1993) reported that the number of English language learners categorized as having limited English proficiency increased by 51.3% between 1985 and 1991.

A profile of U.S. students whose primary language is not English showed that over 65% of these students are in grades K–6, 18% are in middle school, and 14% are in high school. This profile also showed that almost 75% of these students speak Spanish as their native language, followed by Vietnamese (4%), and Hmong, Cantonese, Cambodian, and Korean (2% each). In addition, almost 2.5% speak one of 29 different American Native languages (Navarrete & Gustke, 1996).

Another significant characteristic of English language learners is that a large proportion of them live in high poverty areas. According to a national study of services for students categorized as having limited English proficiency, the socioeconomic status of this population of students is below that of the general school population, as measured by their eligibility for free or reduced-price school lunches (Fleischman & Hopstock, 1993).

In school, the greatest difference between English language learners and their monolingual English-speaking peers is the magnitude of learning expected of them. English language learners, who are all learning a second language, need to work toward English proficiency for both social and academic purposes, face the same academic challenges faced by their monolingual peers, and, to the extent possible, continue development of their native language abilities. Compounding these challenges is the fact that only a small subset of English language learners who come from other countries have strong educational backgrounds (LaCelle-Peterson & Rivera, 1994). For most English language learners, achieving educational success is a daunting task, but the data collected by most states do

not provide enough information to adequately assess the academic standing of these students in the nation's schools (Clements et al., 1992).

How Do Language and Culture Affect the Learning of English Language Learners?

Many factors affect the academic performance of English language learners. Poverty and social inequities block some from achieving success while others simply do not have sufficient access to educational resources and opportunities. However, there is growing evidence that many children do poorly in school mainly because their cultural frames of reference do not match those of the mainstream culture reflected in American classrooms (García & Pearson, 1994; Irvine, 1992). Learning is both a cultural and social process, and students construct knowledge by relating academic content to their lives and by learning from others. Therefore, an English language learner's poor performance in school does not necessarily come from lack of competence in school subjects. Instead, difficulty in school for such a student may be caused by learning tasks that are poorly matched to his or her home culture and the cultural orientations that powerfully influence learning (August & Pease-Alvarez, 1996; Banks & Banks, 1993; Estrin & Nelson-Barber, 1995; Ogbu, 1992).

Culture—defined as a way of life that is shared by members of a population (Ogbu, 1988)—strongly influences what people think is important (values), what they think is true (beliefs), and which behaviors they perceive to be appropriate (norms) (Irvine & York, 1994). How people categorize the world, organize information, and interpret their experiences differ strikingly from culture to culture. "Cultural and linguistic diversity bring with them diversity in cognitive and communicative styles, problem-solving approaches, systems of knowledge, and methods and styles of assessment. What counts as intelligent behavior is variable from

culture to culture; what counts as knowledge and evidence for knowing something, as well as appropriate ways of displaying knowledge are also culturally variable" (Farr & Trumbull, 1997, p.15). Each English language learner brings to the school setting a distinctive set of cultural values, beliefs, and behavioral norms that reflect his or her cultural way of understanding the world. Because of these "cultural differences in ways of knowing and learning," children from varied backgrounds not only speak and interact differently but also think and learn in distinct ways (August & Pease-Alvarez, 1996).

While a child's interactions with family and community influence his or her language use (Shepard, 1992), language and culture also "shape how people conceive of, demonstrate, and measure learning" (Koelsch, Estrin, & Farr, 1995) because cultures vary in their methods of teaching and assessing children in both informal (home and community) and formal (school) settings (Estrin, 1993). As Gardner (1983) observed in outlining his theory of multiple intelligences, while all children develop symbolic competence, they learn quite different symbol systems that reflect the values, beliefs, and norms of their respective cultures. How people use language to structure learning and to show what they have learned varies from culture to culture.

At the same time, learning depends upon language. A language is both the primary medium through which people experience the world and the primary symbol system a culture uses to describe and interpret its environment and to communicate and represent its knowledge. "It is through language that we learn about the world that surrounds us—how to interact with others and objects within that world, how to think about it, how to represent ourselves within it" (Farr & Trumbull, 1997, p. 89). In school, language is used to structure and communicate learning tasks. It also influences how students conceptualize tasks and provides an important tool for understanding and solving problems. It is no surprise, therefore, that

English language learners face unique challenges when they tackle school tasks.

In the daily life of schools, several linguistic and cultural factors affect the ability of English language learners to succeed academically.

- Language use in schools does not match the cognitive and communication patterns of many English language learners' home cultures.
- The speaking patterns used for instruction are not familiar to many English language learners.
- To understand what is expected in a learning situation, these learners need to have prior knowledge of the mainstream culture.
- The English language facility of these students is insufficient for the language demands of many learning tasks.

Language Use in School

Typically, language use in school mirrors the mainstream culture and does not accomodate how students from varied backgrounds use language to learn and demonstrate their learning. Expository styles, patterns of speaking, methods of argumentation, and even rules of good writing differ from one culture to another. For example, one typical European-American style of language use presents a topic in a sequential and linear way, providing evidence and then drawing conclusions. However, Asian and Native American language styles tend to be holistic and circular, presenting multiple topics that are interrelated. In some cultures, it isn't appropriate to ask people questions, while in other cultures students are not accustomed to being asked to respond to a timed task. It is not surprising then that many English language learners experience difficulties when their orientation to language use is so different from that of the schools they attend (Estrin, 1993; Hernandez, 1994; García & Pearson, 1994).

Patterns of Speaking in the Classroom

In classroom learning situations, students have to understand the rules for speaking and the acceptable patterns for communicating what they know.

Adopting discourse patterns that are grounded in the communication styles of the mainstream culture is much more difficult for children learning a second language (Cazden, 1986; August & Hakuta, 1997).

> Mainstream school culture has promoted a widespread discourse pattern for classroom discussions. Interactions tend to follow a pattern in which the teacher initiates an interaction, students respond, and the teacher evaluates. Teachers use this pattern for a variety of classroom discourse functions, including assessing student learning. Inferences that teachers draw from such interactions assume that students are familiar with and recognize the discourse function of the pattern. Sociolinguistic evidence does not support the validity of such an assumption. (García & Pearson, 1994, p. 364)

When English language learners do not respond to the classroom speaking patterns that are unfamiliar to them, educators often mistakenly think that their ability to learn, rather than their communication style, is the source of difficulty.

Language Demands of Learning Tasks

The language of instructional tasks presents a major obstacle to many English language learners because of their limited language skills. The difficulty of a learning task depends to a great extent on the language development and personal experience of the student performing it. So, if a second language learner has to struggle to master lower-level language skills, she or he will be at a disadvantage when responding to tasks which require higher-order language skills. Furthermore, language difficulties are accentuated when a learning task provides little context that makes it clear and meaningful to the student (Farr & Trumbull, 1997).

Prior Cultural Knowledge

Students bring their common experiences and understandings to learning situations. When a learning task draws on familiar cultural knowledge and home and community language uses, a student's prior knowledge helps

him or her understand what is expected. For students who are familiar with mainstream cultural assumptions, continuities exist between home culture and school; for many English language learners these continuities do not exist. Thus, schools that do not make efforts to connect learning activities to the cultural orientations and prior knowledge of English language learners place these students at an educational disadvantage. Because the home and school cultures of many English language learners do not match, their learning potential is underestimated and their strengths are ignored. (Baker & O'Neil, 1995; Saville-Troike, 1991; Koelsch et al., 1995).

Meaningful learning occurs for students when school experiences connect to the ways their culture has taught them to know and understand the world and to use language to acquire and demonstrate knowledge. Furthermore, how students approach a learning task, formulate an argument, or communicate what they have learned affects how they perform and how teachers evaluate them. Therefore, we cannot fully understand why students behave and perform as they do without an awareness of how cultural differences affect student performance (García & Pearson, 1994). While an understanding of a culture's influence on learning is important for all students, it is an essential aspect of addressing the needs of English language learners.

Creating meaningful learning contexts for English language learners involves noticing how instruction and assessment connect to these students' cultural experiences and prior knowledge. To draw on cultural contexts, a teacher must know about different cultures, be aware of the range of language uses across cultures, and understand that difference in communication and thinking style does not mean deficiency in ability. "Teacher awareness of cultural and linguistic variation does not mean that teachers have to come up with different teaching strategies for each student; it simply helps teachers appreciate the range of styles students

bring to the classroom" (Koelsch et al., 1995, p. 18). With a new appreciation of the cultural contexts that their students bring to learning, teachers can find pertinent cultural examples, recreate classroom discourse so that what is being taught connects to what students already know, and "probe the school community, and home environments in a search for insights into students' abilities, preferences, motivations, and learning approaches" (Irvine & York, 1995, p. 494). Villegas (1991) described this approach to teaching diverse student populations as "mutual accommodation," for "both teachers and students adapt their actions to achieve the common goal of academic success with cultural respect" (p. 12).

How Have Assessment Policies Affected The Education of English Language Learners?

Assessment policies that were designed without the diversity of today's population of English language learners in mind exert a powerful influence over every aspect of these students' educations. The policies determine how such students are identified and classified in the school population, what their placement is in the school program, and how their progress is monitored. How schools interpret the performance of English language learners on various tests and assessments influences both teacher beliefs about the abilities of these students and teacher expectations about the kinds of instruction these students should receive. As a result, assessment has compounded the difficulties English language learners face while trying to gain access to the high-quality education they deserve.

> That minority and low-income children often perform poorly on tests is well known. But the fact that they do so because we systematically and willfully expect less from them is not. Most Americans assume that the low achievement of poor and minority children is bound up in the children themselves or their families. 'The children don't try.' 'They have no

> place to study.' 'Their parents don't care.' 'Their culture does not value education.' These and other excuses are regularly offered up to explain the achievement gap that separates poor and minority students from other young Americans. (Commission on Chapter I, 1992, pp. 3-4)

Traditional testing policies and practices have blocked educational opportunity for English language learners for four main reasons:

- Traditional testing has often been culturally biased.
- Tests have often been unable to measure what English language learners actually know.
- Tests have been used for program placement.
- English language learners have often been excluded from national and state assessment programs.

CULTURAL BIAS

As America's schools have grown more diverse, the gaps have widened between the typical achievement scores of students from diverse, non-mainstream populations and those of mainstream English-speaking students. Findings from a recent Congressionally-mandated study using data for first- and third-graders in the 1991-92 school year showed that English language learners lagged behind other elementary school students as measured by grades, retention in grade, and teacher judgements of student ability. This study also showed that English language learners are over-represented among the segment of the student population that scores below the 35th percentile on nationally normed achievement tests (U.S. Department of Education, 1993, 1996). Other studies and national assessments show that, as a group, students who are Hispanic, Native American, African-American, or for whom English is a second language, do not perform as well on formal tests as the population of students that can be categorized as mainstream English-speakers (Mullis & Jenkins, 1990; Educational Testing Service, 1988; National Center for Education Statistics, 1988). Further

analyses of the test results of low-achieving student populations show that poverty and English proficiency are critical factors contributing to low test scores (Pennock-Román, 1992; Rodriguez, 1992).

Discrepancies between the test scores of one group of students and those of another are caused in part by differences in quality of education that result from wide disparities in the financial resources available to schools, unequal access to high-quality curriculum and instruction, educational practices that are aligned with the needs of one group of students but not those of another, staff who are not prepared to teach students from different backgrounds, and discrepancies in parent and community involvement (August & Pease-Alvarez, 1996; Darling-Hammond, 1994; Ferguson, 1991; Oakes & Lipton, 1990; Williams, 1996). However, test bias favoring mainstream student populations is also part of the problem. Content bias and norming bias are two main components of cultural bias in tests (Duran, 1989; Geisinger, 1992).

Content Bias

Content bias occurs when a test's content and procedures reflect the language structure and shared knowledge of the dominant culture or when test items do not include activities, words, or concepts familiar to non-mainstream students (Mercer, 1989; Neill & Medina, 1989; Medina & Neill, 1990). "It is most severe when test tasks, topics, and vocabulary reflect the culture of mainstream society to such an extent that it is difficult to do well on a formal test without being culturally assimilated" (García & Pearson, 1994, p. 344). By including only a limited range of knowledge and ways of expressing knowledge familiar to the mainstream culture, tests whose content is biased have contributed to school policies that exclude students. "Rather than enabling students to bridge the differences between their own backgrounds and the knowledge expected in schools, tests reify the cultural forms and content of knowledge of the dominant groups, and

provide no opportunity for alternate expressions of competence" (Neill, 1995, p. 122).

Norming Bias & Validity

Norming bias occurs when the population samples used to determine whether the content of a test is valid for specific student populations are not representative of minority groups. Because the inferences drawn from test scores are likely to be accurate only for populations for which the test has been validated, assessing English language learners with instruments written in English and normed on monolingual English-speaking students will yield highly questionable results. Inferences made about student competence based on such data are prone to be invalid and can lead to damaging consequences (LaCelle-Peterson & Rivera, 1994). Furthermore, if the time limits set for tests are established during pilot testing with predominantly monolingual samples, time requirements are often set that disadvantage students from backgrounds other than those for which the test was validated. In this situation, English language learners are likely to have too little time to complete the test since it often takes them longer to understand test questions (Mestre, 1984; García & Pearson, 1994).

Because of cultural bias, many tests used in schools merely indicate how familiar students are with mainstream cultural knowledge and ways of demonstrating knowledge (Estrin & Nelson-Barber, 1995). Culturally biased tests do not provide teachers and administrators with adequate tools for assessing culturally varied ways of learning and demonstrating understanding.

INADEQUACY OF ASSESSMENT TOOLS

Though many educators might say that students in America—including English language learners—are over-tested, in many cases current

assessment methods do not provide an adequate picture of what English language learners actually know and can do (LaCelle-Peterson & Rivera, 1994). While testing programs have been limited in the extent to which they adequately measure any student's higher cognitive abilities, they have been particularly limiting for English language learners because tests written in English cannot adequately assess even the content knowledge of these students. Major limitations in assessment tools for English language learners are summarized below.

Difficulties Posed by Unfamiliar English Vocabulary

Unfamiliar English vocabulary poses difficulties for English language learners, and they are at a disadvantage when knowledge of uncommon terms is essential for understanding test instructions, an item, or a passage (Chamot, 1980; García & Pearson, 1991). As the Standards for Educational and Psychological Testing (American Educational Research Association, 1985) point out, whenever students who are still in the process of learning English take tests written in English, regardless of the content or intent of the test, their proficiency in English will also be tested. Because any test written in English is to some degree a test of a student's English language proficiency, such tests may be invalid or unreliable measures of English language learners' academic proficiencies.

Limited Ways of Demonstrating Knowledge

It is particularly important to employ multiple types of assessment when evaluating the ability or achievement of English language learners, but few assessment tools of this type are currently in use. New assessment tools are being designed to allow students to demonstrate knowledge in a variety of different ways. However, even when multiple measures are used, current forms of assessment are not sufficiently responsive to the varied ways of demonstrating culturally-based knowledge (García & Pearson, 1994; Neill, 1995).

Difficulty of Assessing a Variety of Language Skills

Student proficiency in the distinct language skills of reading, writing, speaking, and listening can be difficult to determine. A student's skills may vary considerably from one distinct area of language competence to another. Furthermore, proficiency in a native language may influence both a student's ability to learn English and his or her ability to learn new content in either language (Clements et al., 1992). However, current measures capture neither what English language learners know in each of their languages, nor how these students acquire and use their languages (García, 1992). While research indicates that a student's first language is an important resource that contributes to second language development and improvement of thinking skills (Hakuta, Ferdman, & Diaz 1987), available tests cannot assess learning that draws upon the interactions of two languages.

Few Assessments in Native Languages

Assessment tools are not available in students' native languages. The difficulties of assessing English language learners are compounded because few tools exist for assessing large and established English language learner populations (such as Spanish-speaking students) and virtually none exist for assessing children who speak languages that are less common in the United States (Hernandez, 1994).

USE OF TESTS FOR PROGRAM PLACEMENT

Because schools rely too heavily on test scores for program placement, many have often tracked English language learners into low-ability classrooms in which rote learning and low-cognitive tasks are emphasized and few opportunities exist to practice high-level thinking skills. This use of tests has led to a self-perpetuating cycle in which children who have been sorted according to test scores are placed into educational settings where they receive instruction that focuses on low-level basic skills; then, subse-

quent test results are used to justify initial placements. The type of education provided in these settings ensures that these students will not develop the thinking and analytic skills they need for the future.

In many cases, English language learners have been placed inappropriately in special education classes. Often, based on results of tests that are administered in English and not designed with diverse cultural perspectives in mind, children who use languages other than English are misdiagnosed as having communication disorders. Because of the devastating effects misused assessments have sometimes had on English language learners, educators need to ask themselves, "How can we fairly assess children for possible disabilities when they are not proficient in the language of testing?" (Damico & Hamayan, 1991; Hernandez, 1994).

Because skill with language has a major influence on test performance and testing controls access to educational opportunity, language competency indirectly controls students' opportunities. Sound and accurate tests are needed to assess the academic achievement, diagnose the special educational needs, and predict the academic success of English language learners. However, this will require significant progress in the development and use of unbiased assessment methods and instruments (Lam, 1991).

EXCLUSION FROM ASSESSMENTS

LaCelle-Peterson and Rivera (1994) point out that two predominant options have characterized policies for assessing English language learners. In some cases, English language learners have been tested without considering whether an assessment was technically valid for them as a student population. In other cases, English language learners have been excluded from assessments for a set period of time. LaCelle-Peterson and Rivera conclude that by ignoring validity concerns, testing policies fail to consider the educationally significant differences that distinguish English language

learners from their monolingual peers. On the other hand, exempting English language learners from assessment programs creates a "systemic ignorance" about their educational progress.

Neither our national assessment programs nor most statewide assessment programs have provided adequate data on the academic progress of English language learners. Procedures used prior to 1990 for the National Assessment of Educational Progress (NAEP) allowed schools to exclude a student who was part of the sample population if he or she was categorized as having limited English proficiency and if the local district judged this student incapable of participating meaningfully in the assessment. Beginning in 1990, NAEP defined in greater detail the conditions for excluding students with limited English proficiency, and over two-thirds of students identified as having limited proficiency in English were excluded from NAEP testing in 1992. Upon careful examination, it appears that the exclusion criteria contributed to differences in exclusion rates across states participating in the NAEP Trial State Assessment because of subjective interpretations by local district staff (Olsons & Goldstein, 1997).

Beginning with the 1995 NAEP field test, new procedures were put in place to include a more representative population of students with limited proficiency in English in the assessment sample. Inclusion criteria were revised to promote appropriate and consistent decisions about the inclusion of students with limited English proficiency, and the field test employed various accommodations and adaptations in the mathematics assessment. The findings from the NAEP field test indicated that the new procedures and accommodation strategies would permit inclusion of more students in the national assessment, that new inclusion criteria were not likely to have as pronounced an effect on inclusion as accommodations, and that decisions about how to use the results of students tested with accommodations still needed to be addressed (Olson & Goldstein, 1997).

Separate surveys of state assessment policies and practices conducted in 1994 by the Council of Chief State School Officers (CCSSO) (August & Lara, 1996) and the Center for Equity and Excellence at George Washington University (Rivera, 1995; Rivera et al., 1997) showed that states have not typically included students categorized as having limited English proficiency in their assessment programs.

The CCSSO study noted the following trends:

- Most states exempt students classified as limited English proficient (LEP) from statewide assessments, although 22 of those states require these students to take the assessments within a given period of time after their exemption (usually one to three years). In most instances, the criteria for exemption are based either on the number of years a student categorized as having limited proficiency in English has been in the U.S. or in a bilingual ESL program, or they are based on the student's level of English language proficiency. Some states use multiple criteria that may include language proficiency scores, the number of years in English-speaking classrooms, participation in a program to develop English language proficiency, school achievement, and teacher recommendations.
- Five states reported that they require students categorized as having limited English proficiency to take state assessments. Nevertheless, three of these states also indicated that these students may be exempted under certain conditions.
- Twelve states offer native language assessments—primarily in Spanish. Six of these states reported that the assessments are based on the states' content standards, thus allowing the states to determine Spanish-speaking students' proficiencies on specific learning standards.
- When primary language assessments are not available, many states modify their English language assessments to accommodate students with limited English proficiency. The accommodations sometimes involve altering the administration process or changing the assessment instrument. Some accommodation strategies used by states include translating tests, simplifying directions, administering

assessments orally or in small groups, and allowing students to use dictionaries or take extra time.

- Very few states disaggregate achievement data to examine how student performance relates to LEP status. However, the Improving America's Schools Act (IASA), whose requirements call for the inclusion of all students in statewide assessments and for the disaggregation of assessment results by language status, is likely to change how data are reported (August & Lara, 1996).

Survey findings show that states have had difficulty developing appropriate policies for including students categorized as having limited English proficiency in statewide assessment programs. This suggests the need for research and policy development in several areas. There is a need to refine policies for reporting data about students categorized as having limited English proficiency and to include these students in state accountability reports. The effectiveness of test modifications for English language learners needs to be documented. Also, the implications of using test modifications with "limited English proficient" students who possess different levels of proficiency must be evaluated. Finally, further research is needed to ensure high technical quality in translated tests (Rivera et al., 1997).

Will Assessment Reform Help English Language Learners?

Assessment reform has the potential of improving the quality of learning for English language learners. New approaches to assessment offer varied ways for students to demonstrate what they know and can do while at the same time measuring the learning of all students against high standards. It is likely that the scope of the reform movement will provide the leverage needed to raise expectations for English language learners, and the emphasis on higher level skills should improve the quality of teaching provided to them. However, there are genuine concerns that English language learners

have been kept on the periphery of assessment reform. Some have argued that there is still doubt about whether new assessments will benefit students from poor and culturally diverse backgrounds unless far greater attention is given to equity and opportunity to learn. The influence of linguistic and cultural factors on assessment validity and fairness and the need for greater access to high quality instruction are areas that still need to be addressed. The hopes and cautions of assessment reform for English language learners are discussed below.

RAISING EXPECTATIONS FOR LEARNING

In light of standards-based education reform, the purpose of assessment has been reevaluated. Today, assessments are being used to determine which students are obtaining the knowledge and skills essential to success in today's society so that schools can ensure all students are provided with opportunities to achieve at high levels. Standards-based reform has motivated states to develop assessments that determine whether students have attained high learning standards, and has focused attention on the need to ensure that all students participate in national, state, and local assessments of student progress. Therefore, assessment reform can improve the quality of education provided to English language learners by prompting states, districts, and schools to overcome the limitations that low expectations have placed on the achievement of students who are not yet proficient in English. Standards-based education thus requires "consideration of how assessments, both those currently in use and those which states and school districts are developing, will enable all students, including limited English proficient students, to demonstrate what they know and can do" (Rivera & Vincent, 1996, p. 2).

MORE FLEXIBILITY IN ASSESSMENT

By creating assessment tools that are more adaptable and flexible than those used in traditional testing, assessment reform offers the possibility of using diverse methods to tap the multiple intelligences and talents of students. The work of Kornhaber and Gardner (1993) underscores the importance of looking at the multiple ways students learn and illustrates how student strengths and talents not shown by standardized tests can be illuminated through varied classroom opportunities to demonstrate competence. Because they are designed to reveal more about what students have learned and to take the context of student learning into consideration, new models of assessment promise to be more useful in determining how well public education is serving English language learners (Farr & Trumbull, 1997; Hafner & Ulanoff, 1994; Saville-Troike, 1991). However, including diverse student populations in assessment programs that measure higher order proficiencies requires assessment methods that offer varied paths to demonstrating excellence (Lachat, 1994). By providing English language learners with varied ways of demonstrating what they know and can do, new approaches to assessment can reveal educational "entry points" that might allow educators to build on the strengths of these students and extend their learning into new areas.

IMPROVING TEACHING PRACTICES

Advocates of assessment reform believe that large-scale performance assessment programs will encourage teachers to adopt teaching strategies and classroom activities that encourage thinking and problem-solving (Resnick & Resnick, 1992; Taylor, 1994). At the school level, emerging strategies for assessing student learning through portfolios, exhibitions, projects, and careful observations of children aim to strengthen teaching and learning "by engaging students in more meaningful, integrative, and

challenging work, and by helping teachers to look carefully at student performance to understand how students are learning and thinking" (Darling-Hammond, 1994, p. 21). The fact that these approaches enhance the ability of teachers to look closely at student work may have a positive impact on the quality of teaching provided to English language learners. The variety of methods teachers are being encouraged to use in diagnosing student learning may deepen understanding of different learning styles. However, new forms of assessment will only improve the instruction English language learners receive if these assessments can provide information that is sensitive to both the cultural and individual factors that are relevant to a student's success in school (García & Pearson, 1994).

INCLUDING ENGLISH LANGUAGE LEARNERS IN ASSESSMENT DEVELOPMENT

Though education researchers and advocates have cautioned that "one size fits all" reform won't work and that even the most promising approaches to assessment should not be assumed to work for all populations, assessment reform efforts have neither adequately addressed the needs of English language learners nor fully considered how emerging assessments might affect distinct student populations. "The implicit guiding assumption [of reform] appears to be that whatever curricular revisions and/or assessment innovations contribute to the success of monolingual students will also work for English language learners—that once English language learners know a little English, the new and improved assessments will fit them too" (LaCelle-Peterson & Rivera, 1994, p. 56). An uncritical assumption that all students can be tested in the same ways will likely result in a failure to draw upon students' particular strengths and ways of knowing, will widen the achievement gap between English language learners and the mainstream student population, and will lead to further exclusion of poorly served populations of students. Though having students perform

the same task under the same conditions gives the appearance of equity, meaningful equity would allow all students to put their "best foot forward" and invite students to employ diverse ways of solving problems and accomplishing tasks (García & Pearson, 1994).

Farr and Trumbull (1997) caution that new assessment practices may have limited utility for English language learners because of the common practice of getting reforms in place for "the majority" and then trying to adapt them to "special populations," often after financial and human resources have been exhausted. They argue that assessment development should first focus on special populations instead.

> Most of the interventions have been developed from what was thought to be effective for the general student population. We submit that an approach that focused first on underserved populations would be singularly appropriate for the development of assessments as well. Instead, we have seen a pattern of exempting minority students from assessments or thinking of them after the fact, most notably through the development of a native-language version of an already-developed test for mainstream students. This type of development-by-afterthought will not accomplish the social changes and correction of invalid assumptions that must occur if we are to have a truly equitable educational system—a system that replaces the notions of disadvantage and compensatory education with notions that acknowledge the competence of all students. (Farr & Trumbull, 1997, p. 177)

Because assessment reform does not automatically eradicate test bias, the lack of focus on English language learners in assessment reform has heightened concerns that the greater reliance on language-dependent skills and situational contexts in performance-based assessments may actually increase the sources of cultural bias in emerging assessment programs. Moreover, critical thinking tasks that involve making judgements or expressing values may go against the norms of some cultural groups (Farr & Trumbull, 1997). García and Pearson (1994) point out that experts in multicultural education have shown how difficult it is for mainstream

educators to identify topics that are culturally relevant to minority students and that even the involvement of minority educators in selecting and developing topics, tasks, and rubrics cannot guarantee an assessment's fairness to a particular minority population. Developing performance tasks that can fairly assess students who approach problems from distinct cultural perspectives is a complex challenge with significant implications for whether state and national assessments will be valid and fair for English language learners (Winfield, 1995).

Wolf et al. (1991) point out that the technical demands of test construction—from developing test items to establishing reliability in scoring —are typically given precedence over the need to ensure that tests are both valid and fair for all tested student populations. Even in developing new and innovative assessments, this priority often continues to apply. Admittedly, a few items on any test are likely to be unfair to some students. However, when students whose cultural, language, and economic backgrounds and frames of reference have not been considered by test developers face a disproportionate number of test items that are unfair to them, test results will be invalid for these students. Performance-based assessments, which call for students to use background knowledge and reasoning strategies to make judgements, analyze, and solve problems, only make validity an even more complex issue (Farr & Trumbull, 1997). Therefore, although new assessments offer many potential advantages, "many forms of bias will remain, as the choice of items, responses deemed appropriate, and content deemed important are the product of culturally and contextually determined judgments, as well as the privileging of certain ways of knowing and modes of performance over others" (Darling-Hammond, 1994, p.17). To ensure that current reform efforts create assessments that are valid and fair for English language learners, linguistic and cultural factors must be weighed *during* the assessment development process.

OPPORTUNITY TO LEARN

If current efforts to improve the quality of learning in the nation's schools are to succeed, they must seek to achieve both excellence and equity. Standards-based assessment reform alone cannot create greater educational opportunity for populations of students that in the past have not received high quality educations. Gordon (1992) underscores that assessment reform should not occur in a vacuum, but must consider the complex societal conditions that control access to essential resources.

> There are those of us who are sympathetic to standards and assessment, but insist that it is immoral to begin by measuring outcomes before we have seriously engaged the equitable and sufficient distribution of inputs, that is, opportunities and resources essential to the development of intellect and competence. So we confront the questions of testing in the face of psychometric, pedagogical, political, economic, psychological, cultural, and philosophical problems, and there appear to be few who are prepared to engage such complex problems from these several perspectives. (Gordon, 1992, p. 2)

Winfield (1995) and Darling-Hammond (1994) have also emphasized that if we assume that new forms of assessment will improve teaching practices for students who are poor or from disadvantaged minority populations, we ignore the inadequacy of the instructional conditions that influence the learning of these students.

Many students who are poor or from disadvantaged minority populations have few opportunities to develop the proficiencies reflected in new assessments. They attend schools that receive inadequate funding, have inadequate instructional materials, and have difficulty recruiting highly qualified teachers. At the classroom level, their opportunity to learn is influenced by curriculum content, teacher beliefs, the quality of instruction, time spent on academic tasks, the nature of teacher-student interactions, and the feedback and incentives provided to them (Neill, 1995). Stevens (1996) also adds that such variables as family support,

school climate and environment, and the standards established for student behavior influence a student's opportunity to learn.

For all students to have a fair opportunity to learn the knowledge and skills that are essential for them to participate fully and productively in society, inequities in how our society allocates educational resources must be eliminated. "Without dramatic changes in teaching and resource allocation, minority and LEP students will experience disproportionate failure as they did with the minimum competency tests of the 1970s and 1980s" (Rivera & Vincent, 1996, p. 14). Whether assessment reforms prove helpful to student populations that have been denied access to excellent educations will depend on whether these students gain access to the essential resources and conditions that support learning and achievement.

IMPLICATIONS FOR POLICY AND PRACTICE

Adjust policies and resources so that conditions in schools allow all children, including English language learners, to develop higher-order knowledge and proficiencies.

Policymakers and educational leaders at national, state, and local levels have to address the economic and cultural implications of education reform (Gordon, 1992). Changes must be made in the allocation of resources to schools to ensure all students have access to high quality instruction, skilled teachers, and safe and supportive school environments (Neill, 1995; Stevens, 1996).

Provide English language learners with rich and challenging educational opportunities.

English language learners must be exposed to challenging instruction if they are to achieve at high levels and be fairly assessed by assessment programs designed to measure high level learning. In the past, when tests have been used for program placement, they have failed to ensure that English language learners receive the best education possible. Future policies must

create greater access to rich and varied educational opportunities or a large majority of these students will continue to perform at low levels.

Adopt assessment policies and practices that yield valid information about English language learners.

Cultural bias, fairness, validity, and reliability in the assessment of English language learners must be confronted. Current policies must be evaluated to determine whether they contribute to invalid interpretations of assessment results for English language learners. Policies should require that English language learners are represented in the sample populations used to validate assessments. Translation of assessments also must be validated (Baker & O'Neil, 1995; García & Pearson, 1994; Neill, 1995). To overcome the limitations of current measures, culturally sensitive assessments will need to be developed and sufficient field-testing will need to be conducted to determine whether new assessments are technically adequate and appropriate for English language learners. Many English language learners draw on life experiences that differ from the life experiences of those who develop and administer assessments. These differences must be understood so that these students are not unduly penalized in the assessment process (Winfield, 1995).

Create professional development programs that improve administrator and teacher awareness of how cultural and linguistic factors influence learning.

School administrators and teachers need to become more aware of how culture affects learning. Their leadership is needed in developing school policies and classroom practices that reflect deeper knowledge of different cultures and the belief that cultural differences can be assets rather than obstacles to learning (García & Pearson, 1994; Koelsch et al., 1995). Professional development for administrators and teachers is thus a key to improving how English language learners are taught and assessed.

Encourage teachers of English language learners to make connections between academic tasks and the home cultures of students.

Far more attention must be given to connecting instructional goals, methods, and materials to students' cultural experiences and to the range of learning styles students bring to the classroom (Irvine & York, 1995; Saville-Troike, 1991). This does not mean that learning should be limited to topics that relate to the experiences students bring to school, but rather that these experiences serve as starting points for making knowledge meaningful. The challenge is to make effective instructional use of the personal and cultural knowledge of students while at the same time helping them reach beyond their cultural boundaries (Banks & Banks, 1993).

Give English language learners additional time and support when they are learning classroom uses of language that are unfamiliar to them.

While providing appropriate content helps to facilitate learning for students from non-mainstream backgrounds, helping students to understand how to use language in learning situations is equally important (Estrin, 1993; Hernandez, 1994). Because students' backgrounds influence how they use language, students who learn different patterns of language use at home than those commonly employed in mainstream classrooms should be given instructional support that helps them expand their repertoire of language use (Farr & Trumbull, 1997).

Develop varied approaches to assessment and clear guidelines for interpreting results so that English language learners are not placed inappropriately in special education classes.

Assessments need to distinguish between English language learners who are performing unsatisfactorily in school because of limited exposure to English and children who demonstrate communication disorders and need special education intervention (Hernandez, 1994). If we want to create assessments that do not contribute to the misdiagnosis of student knowl-

edge and skills, we need to develop more dynamic and flexible approaches to assessment that seek to determine what children are capable of learning, not just what they already know (Duran, 1989). For assessments to become fair and valid for a wider range of students, those conducting the assessments need a deeper understanding of the cultural, linguistic, and experiential backgrounds that children bring to learning (August & Hakuta, 1997).

References

American Educational Research Association, American Psychological Association, National Council of Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

August, D., & Lara, J. (1996). *Systemic reform and limited English proficient students*. Washington, DC: Council of Chief State School Officers.

August, D., & Pease-Alvarez, L. (1996). *Attributes of effective programs and classrooms serving English language learners*. The National Center for Research on Cultural Diversity and Second Language Learning.

August, D., & Hakuta, K. (Eds.) (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.

Baker, E., & O'Neil, H. (1995). Diversity, assessment, and equity in educational reform. In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment*. Boston: Kluwer Academic Publishers.

Banks, J., & Banks, C. M. (Eds.) (1993). *Multicultural education: Issues and perspectives* (2nd ed.). Boston: Allyn & Bacon.

Cazden, C. (1986). Classroom discourse. In M. Wittrock, (Ed.), *Handbook of research on teaching* (pp. 432-463). New York: Free Press.

Chamot, A. (1980). Recent research on second language reading. *National Association of Bilingual Education (NABE) Forum*, 3-4.

Cheung, O., Clements, B., & Mieu, Y. (1994). *The feasibility of collecting comparable national statistics about students with limited proficiency*. Washington, DC: Council of Chief State School Officers.

Clements, B., Lara, J., & Cheung, O. (1992) *Recommendations for improving the assessment and monitoring of students with limited English proficiency*. Washington, DC: Council of Chief State School Officers.

Commission on Chapter I. (1992). *High performance schools: No exceptions, no excuses*. Washington, DC: Author.

Damico, J., & Hamayan, E. (1991). Implementing assessment in the real world. In E. Hamayan & J. Damico (Eds.), *Limiting bias in the assessment of bilingual students*. Austin, TX: Pro-Ed.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*, 5-30.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass Publishers.

Duran, R. P. (1989). Assessment and instruction of at-risk Hispanic students. *Exceptional Children, 56*, 2, 154-158.

Educational Testing Service. (1988). *A summary of data collected from Graduate Record Examinations test takers during 1986-1987, Data Summary Report #12*. Newark, NJ: Author.

Estrin, E. (1993). *Alternative assessment: Issues in language, culture, and equity*. San Francisco: Far West Laboratory.

Estrin, E., & Nelson-Barber, S. (1995). *Issues in cross-cultural assessment: American Indian and Alaska Native students*. (Far West Laboratory Knowledge Brief #12). San Francisco: Far West Laboratory.

Farr, B., & Trumbull, E. (1997). *Assessment alternatives for diverse classrooms*. Norwood, MA: Christopher-Gordon.

Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28* (2), 465-498.

Fleishman. H., & Hopstock, P. (1993) *Descriptive study of services to limited English proficient students, volume I, summary of findings and conclusions*. Prepared for U.S. Department of Education. Arlington, VA: Development Associates, Inc.

García, G., & Pearson, P. (1991). The role of assessment in a diverse society. In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 253-278). New York: Teachers College Press.

García, G. (1992). *The literacy assessment of second-language learners*. University of Illinois at Urbana-Champaign, Center for the Study of Reading, Technical Report No 559.

García, G., & Pearson, P. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of Research in Education, 20*. Washington, DC: American Educational Research Association.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Geisinger, K. F. (Ed.). (1992). Fairness and psychometric issues. In K.F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 17-42). Washington, DC: American Psychological Association.

Glaser, R. (1987). The integration of instruction and testing: Implications from the study of human cognition. In D.C. Berliner and B.V. Rosenshine (Eds.), *Talks to teachers*, (pp. 329-341). New York: Random House.

Gordon, E. (1992). *Implications of diversity in human characteristics for authentic assessment*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Hafner, A., & Ulanoff, S. (1994). Validity issues and concerns for assessing English learners: One district's approach. *Education and Urban Society, 26*, 367-389.

Hakuta, K., Ferdman, B., & Diaz, R. (1987). Bilingualism and cognitive development: Three perspectives. In Rosenberg, S. (Ed.), *Advances in applied psycholinguistics*, Vol. 2.

Henderson, A., Donley, B., & Strang, W. (1995). *Summary of the bilingual education state educational agency program survey of states' limited English proficient persons and available educational resources 1993-94*. Arlington, VA: Development Associates, Inc.

Hernandez, R. (1994). Reducing bias in the assessment of culturally and linguistically diverse populations. *Journal of Educational Issues of Language Minority Students, 14*, 269.

Irvine, J. J. (1992). Making teacher education culturally responsive. In M.E. Dilworth (Ed.), *Diversity in teacher education* (pp. 79-92). San Francisco: Jossey-Bass Publishers.

Irvine, J. J., & York, D. E. (1995). Learning styles and culturally diverse students: A literature review. In J. Banks & C. Banks (Eds.), *Handbook of research on multicultural education* (pp. 484-497). Boston: Allyn & Bacon.

Koelsch, N., Estrin, E., & Farr, B. (1995). *Guide to developing equitable performance assessments*. San Francisco: WestEd.

Kornhaber, M., & Gardner, H. (1993). *Varieties of excellence: Identifying and assessing children's talents*. New York: Columbia University, Teachers College, National Center for Restructuring Education, Schools, and Teaching.

Lachat, M. A. (1994). *High standards for all students: Opportunities and challenges*. South Hampton, NH: Center for Resource Management, Inc.

LaCelle-Peterson, M., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review, 64* (1), 55-75.

Lam, T. (1991). Testing of limited English proficient children. In Kathy E. Green (Ed.), *Educational testing: Issues and applications*. New York: Garland Publishing Company.

Medina, N., & Neill, D. M. (1990). *Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in America's public schools* (3rd ed.) Cambridge, MA: FairTest.

Mercer, J. (1989). Alternative paradigms for assessment in a pluralistic society. In J.A. Banks & C.A.M. Banks (Eds.), *Multicultural education: Issues and perspectives* (pp. 289-304). Boston: Allyn & Bacon.

Mestre, J. (1984). The problem with problems: Hispanic students and math. *Bilingual Journal, 32*, 15-19.

Mullis, I. V. S., & Jenkins, L. B. (1990). *The reading report card, 1971-88: Trends from the nation's report card*. Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service.

National Center for Education Statistics. (1988). *Education indicators*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Navarrete, C., & Gustke, C. (1996). *A guide to performance assessment for linguistically diverse students*. Albuquerque, NM: Evaluation Assistance Center-West.

Neill, M. (1995). Some prerequisites for the establishment of equitable, inclusive multicultural assessment systems. In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment*. Boston: Kluwer Academic Publishers.

Neill, M. & Medina, N. J. (1989). Standardized testing: Harmful to educational health. *Phi Delta Kappan, 70*, 688-697.

Oakes J., & Lipton, M. (1990). *Making the best of schools*. New Haven, CT: Yale University Press.

Ogbu, J. U. (1988). Cultural diversity and human development. In D.T. Slaughter (Ed.), *Black children and poverty: A developmental perspective* (pp. 11-28). San Francisco: Jossey-Bass Publishers.

Ogbu, J. U. (1992). Understanding cultural diversity and learning. *Educational Researcher, 21* (8), 5-14.

Olsen, R. E. W. B. (1993). A survey of LEP and adult ESL student enrollments in US public schools. In *Language minority student enrollment data*. Symposium conducted at the TESOL Convention, Atlanta, GA.

Olson, J. R., & Goldstein, A. A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress*. National Center for Education Statistics, U.S. Department of Education, Washington, DC.

Pennock-Román, M. (1992). Interpreting test performance in selective admissions for Hispanic students. In K.F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 99-135). Washington, DC: American Psychological Association.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for education reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Rivera, C. (1995). *How we can ensure equity in statewide assessment programs. Findings from a national survey of assessment directors on statewide assessment policies for LEP students*. Paper presented at annual meeting of the National Conference on Large-Scale Assessment, June 18, 1995, Phoenix, AZ. Washington, DC: George Washington University Institute for Equity and Excellence in Education and Evaluation Assistance Center-East.

Rivera, C. & Vincent, C. (1996). *High school graduation testing: Policies and practices in the assessment of limited English proficient students*. Paper presented at annual meeting of the National Conference on Large-Scale Assessment, June 24, 1996, Phoenix, AZ. Washington, DC: George Washington University Center for Equity and Excellence in Education.

Rivera, C., Hafner, A., & LaCelle-Peterson, M. (1997). *Statewide assessment programs: Policies and practices for the inclusion of limited English proficient students*. ERIC Clearinghouse on Assessment and Evaluation. Washington, DC: The Catholic University of America.

Rodriguez , O. (1992). Introduction to technical and societal issues in the psychological testing of Hispanics. In K.F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 11-15). Washington, DC: American Psychological Association.

Saville-Troike, M. (1991). *Teaching and testing for academic achievement: The role of language development.* National Clearinghouse for Bilingual Education FOCUS: Occasional Papers in Bilingual Education, No. 4. Washington, DC: NCBE.

Shepard, L. A. (1992). Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In B. Gifford & M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Sternberg, R. (1985). *Beyond IQ.* New York: Cambridge University Press.

Stevens, F. I. (1996). Closing the achievement gap: Opportunity to learn, standards, and assessment. In B. Williams (Ed.), *Closing the achievement gap: A vision for changing beliefs and practices* (pp. 7-95). Alexandria, VA: Association for Supervision and Curriculum Development.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal, 31,* 231-262.

U.S. Department of Education. (1993). *PROSPECTS: The congressionally mandated study of educational growth and opportunity, interim report: Language minority and limited English proficient students.* Washington, DC: Author.

U.S. Department of Education. (1996). *PROSPECTS: The congressionally mandated study of educational growth and opportunity, first annual report.* Washington, DC: Author.

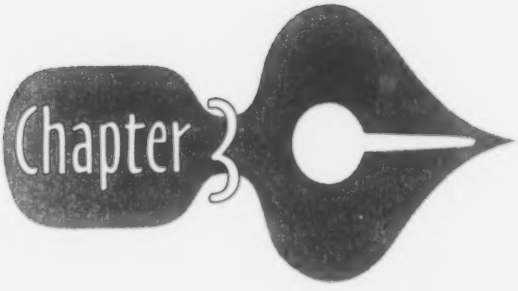
Waggoner, D. (Ed.). (1992, July, September). *Numbers and needs: Ethnic and linguistic minorities in the United States.* (Available from Numbers and Needs, Box GIH/B, 3900 Watson Place, NW, Washington, DC 20016).

Williams, B. (Ed.). (1996). *Closing the achievement gap: A vision for changing beliefs and practices.* Alexandria, VA: Association for Supervision and Curriculum Development.

Winfield, L. (1995). Performance-based assessments: Contributor or detractor to equity? In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of Research in Education.* Washington, DC: American Educational Research Association.

BEST COPY AVAILABLE


Chapter 3

Inclusive And Equitable Assessment for English Language Learners

Including English language learners in the movement to raise standards of learning for all students will not yield positive results without addressing equity and fairness issues in large-scale assessments and in the practices used to assess ELLs in local school districts.

> Tests that do not accommodate crucial differences between groups of children are inherently inequitable. They do not give all children a fair chance to succeed because they assume that all children come to the testing situation with roughly the same experiences, experiences that are crucial for success. (Meisels, Dorfman, and Steele, 1995)

What Factors Must Be Considered In Order To Assess English Language Learners Equitably?

Achieving fairness and equity in assessment for the increasingly large population of English language learners in America's schools is one of the most challenging aspects of assessment reform. Raising standards of learning for these students means that how they have been assessed in large-scale assessments and at the school district level must be changed. In writing about how assessment can

better serve educational reform for students from diverse cultures, Malcom (1991) proposed several essential conditions that capture the essence of fairness and equity in assessment.

- Rules about what is to be known must be clear to all.
- Ways of demonstrating knowledge must be many and varied.
- Knowledge valued by different groups must be reflected in what we expect children to know.
- Resources needed to achieve must be available to all.

At the heart of equity in assessment is whether the design of new assessments can be responsive to diversity and whether all children will be given adequate preparation in the proficiencies assessed. Several factors affect equity in assessment for English language learners. These include what prior knowledge and language skills assessment tasks require, whether test content, procedures, or scoring criteria are biased, whether tests are valid, and whether all students have the opportunity to learn the material assessed. Each of these factors present a score of issues yet to be resolved. In addition, these factors are interrelated and influence one another.

QUESTIONS ABOUT ASSESSMENT EQUITY

The key issues that policymakers and administrators should consider under these areas have been identified by several researchers who have written extensively on equity issues in assessment for diverse student populations. Questions that can help educators evaluate the equity of assessments are identified below.

Relevant Prior Knowledge

What common experiences and understandings must students have to make sense of the assessment task and solve it?

Can students connect their cultural background and experiences to what is expected in the task?

What information is essential for successful performance?

Will all groups be motivated by the topics provided?

Are the criteria for performance known and familiar to all students—do all students understand what kind of evidence of learning will be valued when the assessment is scored?

(Baker and O'Neil, 1995; Farr & Trumbull, 1997; Saville-Troike, 1991)

Language Demands and Content Bias

What language demands do the tasks—particularly those emphasizing higher-order thinking skills—place on students with backgrounds in languages other than English?

If the task is not primarily meant to assess language facility, what alternative options for displaying understanding are available to students with limited English proficiency?

Are the concepts, vocabulary, and activities important to the assessment tasks familiar to all students to be tested, regardless of their cultural backgrounds?

Is the range of knowledge and ways of expressing knowledge called for in the assessment familiar only to the mainstream culture?

Are the limited topics used in performance assessments relevant to students with many different backgrounds?

(Baker & O'Neil, 1995; Estrin & Nelson-Barber, 1995; Farr & Trumbull, 1997; García & Pearson, 1991, 1994; Medina & Neill, 1990; Neill, 1995)

Validity

Is the test valid for the school populations being assessed?

Has the assessment been validated with culturally and linguistically diverse student populations?

Does the assessment take into account the cultural backgrounds of the students taking the test?

Have all test translations been validated and normed?

(Farr & Trumbull, 1997; LaCelle-Peterson & Rivera, 1994; Rivera & Vincent, 1996)

Procedural Bias and Scoring Criteria

Do assessments unduly penalize students for whom the testing format is unfamiliar or the prescribed time limitations are inadequate because of unfamiliarity with the test language?

Are English language learners given sufficient time to complete an assessment?

Do language differences, cultural attitudes toward test-taking, lack of test-wiseness, or test anxiety unduly penalize some students?

What accommodations would be necessary to give English language learners the same opportunity as monolingual students to demonstrate what they know and can do?

Are the scoring criteria used to judge student performance biased toward the mainstream culture? Are the criteria specific enough to overcome the potential for bias when multiple raters are used to judge the performance of a group of students?

Do scoring criteria for content-area assessments focus on the knowledge, skills, and abilities being tested and not on the quality of the language in which the response is expressed?

Are those scoring the assessment sufficiently familiar with students' cultural and linguistic backgrounds to interpret student performances appropriately and to recognize and score English language learners' responses?

Do those scoring students' work include educators from the same linguistic and cultural backgrounds as the students tested?

(Baker & O'Neil, 1995; Farr & Trumbull, 1997; García & Pearson, 1994; LaCelle-Peterson & Rivera, 1994)

Opportunity to Learn

Have all students had the opportunity to learn the assessed material and to prepare adequately to respond to the assessment tasks?

Have English language learners been placed in challenging learning situations that are organized around a full range of educational outcomes?

Have all students been taught by teachers of equal quality, training, and experience?

What educational resources are available to students? Are comparable books, materials, technology, and other educational supports available to all groups to be tested?

(Baker & O'Neil, 1995; Darling-Hammond, 1994; LaCelle-Peterson & Rivera, 1994)

Will Performance Assessment Benefit English Language Learners?

Alternative assessments, called "performance" or "authentic" assessments, invite students to apply their knowledge to real-world tasks. While the term "performance assessment" indicates that students are asked to demonstrate—through their performances on assessment tasks—that they can apply learned skills and competencies, the term "authentic" suggests that students are asked to perform assessment tasks in practical or "real life" contexts. Thus, authentic assessment can be thought of as a subset of performance assessment (Meisels, et al., 1995). Performance assessments gather evidence by employing many different types of assessment tools, such as oral presentations, exhibitions, portfolios of student work, experiments, cooperative group work, research projects, student journals, anecdotal records, notes from teacher observations, and teacher-student conferencing. Therefore, performance assessments draw on a wider range of evidence than do other forms of assessment.

New forms of assessment offer greater promise of accommodating diversity and improving equity in education than do traditional assessments, but not much research has been done on the use of performance-based assessments with students from diverse cultural, linguistic, and economic backgrounds. Some have cautioned that there might be potential problems in using performance assessments fairly with culturally and linguistically diverse groups of student (García & Pearson, 1994). The potential effects of performance assessment on English language learners are discussed below.

ADVANTAGES OF USING PERFORMANCE ASSESSMENTS WITH ENGLISH LANGUAGE LEARNERS

In summarizing the striking contrasts between performance assessments and group-administered standardized achievement tests, Meisels et al. (1995) highlighted the following valuable features of performance assessments.

Performance Assessments:

- Actively involve both students and teachers in the learning process
- Minimize the likelihood of drawing conclusions from limited performance opportunities
- Offer children from different backgrounds varied ways to display their knowledge and abilities
- Provide information that can be used to form a profile of a student's individual strengths and weaknesses
- Allow teachers to monitor student progress over time and influence ongoing learning

"Rather than generalizing from a narrow task to a larger domain, performance assessment aims to document the broad-based process of learning. The purpose is to follow children's development over time, within and across domains, to create differentiated profiles or portraits of children's accomplishments and repertoires" (Meisels et al., p. 251). For English language learners, performance assessments have several advantages. They are closely linked with instruction, reveal more meaningful information about student knowledge and abilities, and allow students to display competencies in a wide variety of ways.

Closer Link to Instruction

Performance assessments can benefit English language learners when they are embedded in a sound, learner-centered curriculum. Typically, performance-based assessment strategies are integrated with instruction and encourage teachers and students to collaborate in the learning process.

When performance assessments are used to support standards-based curricula, they can benefit English language learners by exposing them to essential knowledge and allowing them to apply it to meaningful situations (Baker & O'Neil, 1995; García & Pearson, 1994; Valdez-Pierce & O'Malley, 1992). In addition, the use of performance assessments may encourage teachers not only to set challenging standards for English language learners, but also to use the information about student learning to adapt instruction to individual students more effectively (García & Pearson, 1994).

Performance tasks invite English language learners to solve real problems and provide them with more control over their learning. Because social context plays an important role in performance assessment, not only are students' experiences seen as relevant, but they are viewed as an essential part of the learning process. The notion of embedding instruction and assessment in a social context is important to the learning of students who have to demonstrate content knowledge through an emerging second language (Hafner & Ulanoff, 1994). If instruction and assessment are connected to meaningful contexts, English language learners will be better able to demonstrate what they know and can do.

More Meaningful Information About Student Knowledge and Abilities

By making greater cognitive demands on students than traditional tests, performance tasks invite a fuller range of responses, provide a richer picture of what students have learned, and allow for the ongoing assessment of higher-order thinking skills (Farr & Trumbull, 1997). Because performance assessments allow teachers to observe the development of student thinking and organizational skills, they can be used to create profiles of the educational progress of English language learners. In fact, research has shown that teachers who use authentic classroom assessment tend to

document the growth of individual students over time and often record their findings in narrative or descriptive formats that can be shared with students and parents (Calfee & Perfumo, 1993; García & Pearson, 1994). In addition, performance assessments allow for cultural adaptations and openly invite student performances that may reflect diverse cultural perspectives. Thus individual teachers may find better ways to document information they regard as important to understanding the learning of students from a variety of different language backgrounds. As García and Pearson noted in 1994:

> For example, in Spanish-English bilingual classrooms, teachers will want to know what literacy tasks a child can complete in English, in Spanish, or in both languages (García, 1992). They will want to know the extent to which their students interpret material and vocabulary from cultural and linguistic perspectives based on their backgrounds or from a mainstream perspective (García, 1991). Similarly, they will want to know the extent to which bilingual students can use their knowledge of native-language reading to help in their second-language reading (Downing, 1984; Jiménez, 1992; Jiménez, García, & Pearson, 1991). Teachers working with dialect-speaking African-American youths on improving their writing also might want to evaluate these students' use of dialect apart from their ability to develop a persuasive essay in standard written English (García & Pearson, 1991). It is difficult to imagine formal assessments that could or would attempt to gather such information. (p. 363)

Wider Range of Ways to Display Competencies

Because performance assessments involve the use of multiple measures, they invite students to draw on multiple intelligences and to display varied cognitive and communicative styles. As a result, these assessments provide a wider range of opportunities for English language learners to show what they know and can do in both language and content areas (Estrin & Nelson-Barber, 1995; Navarrete & Gustke, 1996). At the same time, the flexibility of performance assessment allows teachers to vary the

assessment methods used in order to more accurately diagnose the learning of students whose cognitive and cultural styles may cause them to perform poorly on conventional tests. By offering a range of contexts—including opportunities to work alone, in pairs, or in groups—teachers can vary assessment settings to reflect cultural preferences and also evaluate the impact of these contexts on particular students' progress (García & Pearson, 1994). Because performance assessment allows students to show how they solve tasks, teachers may be able to differentiate between learning problems caused by limited English skills and those caused by limited content knowledge. Also, performance assessments may provide more valid information about a student's developing knowledge (Farr & Trumbull, 1997).

When using performance assessments, an approach called "dynamic assessment" can be employed that helps teachers determine which tasks students can complete independently and which they can complete with varying levels of assistance. "[Dynamic assessment] assumes the stance that assessment should be directed toward finding out what the student is capable of learning (working in the 'zone of proximal development') with the assistance of the teacher rather than toward finding out what he already knows" (Farr & Trumbull, 1997, p. 235). Therefore, dynamic assessment allows teachers to document the progress that students who are learning a second language are making with and without support (García, 1991, 1992).

> Within the philosophical parameters of dynamic assessment, teachers would be able to provide students with background knowledge essential to text comprehension, translate obscure English vocabulary that might block an otherwise transparent linguistic translation, or provide other forms of assistance that bilingual students might need in order to comprehend and complete tasks in English. (García & Pearson, 1994, p. 370)

Designed to reveal how a child learns, dynamic assessment procedures can provide students with a series of increasingly challenging tasks and offer varying levels of assistance to help students perform successfully. Proponents believe that dynamic assessment offers the opportunity to gain insights into how a wide range of children learn, which instructional strategies facilitate learning, and which learners respond best to specific types of instruction. As a result, dynamic assessment can provide information about potentially effective techniques of educational intervention (Farr & Trumbull, 1997).

PROBLEMS OF USING PERFORMANCE ASSESSMENT WITH ENGLISH LANGUAGE LEARNERS

Researchers have noted that there may be disadvantages to using performance assessment with English language learners. Because of the particular ways in which performance assessments are structured, scored, and administered, English language learners could encounter difficulties that would make the assessments unfair to them. What follows is a summary of the main concerns that have been identified about the use of performance assessment with English language learners.

Language and Cultural Demands of Performance Assessments

Performance assessments are particularly demanding for English language learners because they rely heavily on language skills. Because performance assessments require students to read and write more when solving problems and demonstrating their critical thinking, the language demands are greater than those of traditional standardized tests. Even in mathematics and science, students are expected to write explanations of how they went about solving problems. If an English language learner's literacy skills interfere with his or her ability to successfully accomplish an assessment task, it becomes impossible to distinguish between the student's literacy and subject-matter knowledge and skills, and the assessment will provide

little useful information about the student's subject-matter performance (Koelsch, Estrin, & Farr, 1995; Navarrete & Gutsky, 1996).

Consideration also must be given to whether performance assessments are based on culturally specific contexts that are unfamiliar to students from certain cultures. A "real-life" problem may be very real to one student but totally unfamiliar to a student from a different culture. In addition, different cultures have different ways of solving problems and different ways of expressing solutions. For example, an assessment might invite students to make judgments or express values, but these particular forms of demonstrating knowledge may not be compatible with the cultural styles with which a student is most familiar.

For performance assessments to be fair to English language learners, they must take into account how these students use language and provide them with a sufficient context for understanding and responding appropriately to the assessment task. If these considerations are not addressed, then English language learners will perform no better on performance assessments than they do on current traditional academic achievement tests (Navarrete & Gutsky, 1996).

Teacher Bias

Teacher bias is a potential problem in the use of performance assessments with English language learners. Teacher beliefs about new forms of assessment, expectations for students, training in the use of alternative assessments, views about the use of results, and methods of motivating students are likely to affect the performance assessment results of English language learners (Rueda & García, 1992). Cultural bias will not be eliminated just because performance assessments take the place of standardized tests. As experts in multicultural education have pointed out, it is difficult for teachers from mainstream backgrounds to identify topics that are relevant to culturally diverse groups (Banks & Banks, 1993;

87

Hernandez, 1989). Also, when teachers do not consider how students' cultural backgrounds affect their ways of working on a task, they tend to form expectations about how a task will be completed that lead to false impressions about student abilities (García & Pearson, 1994). "Because teachers are typically not trained for, or systematic in their use of performance assessment, they may form impressions of students too quickly and use the data they collect from students to maintain those impressions throughout the year" (Meisels et al., 1995, p. 250). To use performance assessments fairly in a classroom with students from diverse cultural and language backgrounds, teachers must become knowledgeable both about the subject matter being assessed and about students' cultures and languages (García & Pearson, 1991).

Validity, Reliability, and Procedural Issues

Performance assessments are quite different from standardized tests, and their procedures introduce new possibilities for inequities. While some of these potential inequities are connected to the nature of performance assessment tasks, other potential inequities emerge from biases in how performance assessments are scored or from the contexts in which assessments are administered.

Fewer tasks and longer reading passages

When large-scale assessment programs employ the more complex tasks used in performance assessments, fewer topics can be surveyed by the test. Questions have been raised about how this limited number of topics will affect the scores of diverse student populations. When an assessment draws on a limited number of assessment tasks, it increases the likelihood that some children may have had little exposure to the limited content reflected in the assessments (Estrin & Nelson-Barber, 1995). A limited range of topics may not provide adequate opportunities to assess the performance of varied student populations. Some have suggested that these kinds of assessments may produce results that show more about

BEST COPY AVAILABLE

students' mainstream cultural experiences than about actual competencies in the subjects assessed.

Performance assessments often present students with fewer tasks than traditional standardized testing by asking students to respond to a few longer passages rather than a wide range of short reading passages. By using longer passages, performance assessments seek to provide students with more meaningful and authentic opportunities to demonstrate what they know and can do. However, longer passages may be particularly difficult for English language learners, and judgments about the proficiency levels of these students that are based on a more limited sampling of tasks or passages may lead to incorrect inferences about their actual capabilities (García & Pearson, 1994).

Scoring rubrics and bias.

Whether the results of performance assessments of English language learners are reliable or not will depend on how scoring rubrics are developed and how much bias affects the scoring of assessments. Those scoring student responses to performance assessments are expected to apply clearly defined performance criteria to make a sound judgement about the level of proficiency demonstrated. However, even in large-scale performance assessments, only a small number of teachers participate in the design of scoring rubrics. Therefore, the reliability of performance criteria can be undermined if those who design scoring rubrics are not knowledgeable about how to teach and assess children from a variety of cultural and linguistic backgrounds (Baker & O'Neil, 1995).

Furthermore, although research has shown that well-trained raters working with well-defined and articulated scoring criteria can reach high levels of agreement with one another (Shavelson, Baxter, & Pine, 1992), ratings of student performance can be subject to scorer biases based on observable attributes such as a student's ethnicity and gender. "Even when

relatively structured rubrics are used, there is some evidence that raters rate members of their own race or ethnicity higher than those of other races and ethnicities" (Baker & O'Neill, 1995, p. 73). In addition, when student performance on a demonstration or exhibition is assessed, it can be heavily influenced by scorer response to students' verbal skills, dialect, or accent. In many cases, individuals who speak in dialects or with accents are more likely to be judged as less intelligent and less capable" (García & Pearson, 1994, p. 228).

Assessment administration

How performance assessments are administered is likely to vary from one school or classroom to the next, and "differences in procedures such as task directions, the provision of help, and the availability of resources can be counted on to have known and measurable effects on student results" (Baker & O'Neil, 1995, p. 72). Therefore, the context in which a performance-based assessment is administered will affect the validity of the results for English language learners (Winfield, 1995).

What Policies and Practices Should be Followed When Including English Language Learners in Statewide Assessment Programs?

Because adequate resources have never been devoted to addressing the issues of assessment for non-English speakers in America's schools, there are many more questions than answers about the policies and practices that should be followed in including English language learners in large-scale statewide assessment programs (August & Hakuta, 1993; Olson & Goldstein, 1997). However, while the knowledge base is limited, many studies are underway, and core questions are being answered. Proceedings from a national conference on "Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of

Educational Progress (NAEP)" underscored the importance of developing a coherent framework for inclusion, citing three overall principles which are also relevant to state assessment systems (August & McArthur, 1996).

BASIC PRINCIPLES

Maximum Inclusion

Assessment results should represent all students. Every student, regardless of language characteristics, should be included in the assessment population.

Continuum of Strategies

Because no single strategy will enable all English language learners to participate fairly in large-scale assessment programs, a continuum of options should be available to support the participation of these students. These options may include both those that have been proven to be effective as well as untested options that still need to be field-tested.

Researchers suggest that assessment programs should draw on available options and attempt to maximize the number of students who are offered options on the tested or proven end of the continuum. At the same time the feasibility and impact of untested options should be investigated. Using the entire range of options would allow the inclusion of most students, even though "some of the students would only be included through the use of non-comparable assessment strategies." (August & McArthur, 1996, p. 9)

Practicality

Assessments designed to meet the needs of English language learners must be evaluated for their costs, their benefits, their consequences, and the feasibility of their administration. For example, since it may not be feasible to develop native language assessments because of the costs and

psychometric problems involved in getting an equivalent translation of a test from one language to another, other ways of including English language learners who are not proficient in English in assessment programs would need to be explored. Alternative assessment strategies must also take into account whether the requirements and burdens of assessment administration are manageable at the local level and whether the toll of assessment on individual test takers might be too great.

GUIDELINES FOR INCLUDING ENGLISH LANGUAGE LEARNERS IN STATEWIDE ASSESSMENT PROGRAMS

Based on the three fundamental principles above, current research suggests that the following guidelines should be followed in developing and implementing assessment policies and practices that include English language learners in statewide assessment programs to the fullest extent possible.

Consider how to include populations such as English language learners when the assessments are being developed.

Are statewide assessments appropriate, valid, and reliable for English language learners? All too often, states develop and field-test new assessments for the general population, allowing the technical demands of test construction to postpone consideration of whether these new assessments are appropriate and fair for English language learners. Once developed, tests are then reviewed to determine whether a native-language version or some type of accommodation would facilitate the participation of English language learners. However, addressing the needs of English language learners as an afterthought makes it more difficult to develop assessments that are inclusive, valid, and reliable for this population. Instead of adjusting assessments to English language learners after their development, those who specialize in working with English language learners should be asked to participate from the beginning when assessment

policies, items or tasks, and procedures are being developed. (Farr & Trumbull, 1997; Olson & Goldstein, 1997).

Choose assessment content that is appropriate for the diverse populations taking the test.

Both the diverse cultural backgrounds represented in a student population and the amount of knowledge of mainstream culture needed to understand and respond to an assessment item should be considered when developing assessments. Because many English language learners draw on life experiences that differ from those who develop assessments, these students often respond to performance assessments in unanticipated ways. For performance assessments to be fair to English language learners, assessment tasks should be developed with their cultural perspectives in mind (Winfield, 1995).

Field-test assessments with English language learners to ensure validity.

Only assessments that have included English language learners in their field test population samples will be valid for use with these students. Making inferences about the competencies of English language learners from assessments that have been validated with monolingual English-speaking students constitutes an invalid use of assessment data (LaCelle-Peterson & Rivera, 1994). A "best practice" approach in the development of assessment instruments and procedures is to field-test them with a student sample that is representative of all the types of students who will take the assessment (Olson & Goldstein, 1997).

Establish scoring criteria appropriate for evaluating the work of English language learners and train those who score assessments properly.

Assessment scoring criteria must make it possible to determine the content-area knowledge, skills, and abilities being tested while not becoming skewed by the linguistic skill with which student responses are expressed

(LaCelle-Peterson & Rivera, 1994). Otherwise, English language learners will be penalized inappropriately for lacking English language skills. In the case of performance assessments, individuals who are knowledgeable about the cultural and linguistic characteristics of the students being assessed should participate in the development of rubrics for scoring student work. Furthermore, assessment personnel who score the responses of English language learners must be carefully selected and trained.

Field-test assessment options that will maximize the inclusion of English language learners in state assessments.

Large-scale assessments typically employ two options in order to maximize the participation of English language learners:

- *alternative assessments* that modify the assessment instrument to make it easier for students with limited proficiency in English to comprehend
- *accommodations* that adjust test administration procedures to support students with limited proficiency in English

Questions are often raised about whether the results of an alternative assessment are comparable to those of the assessment it replaces, whether alternative assessments offer valid measures of the content being assessed, and whether scoring rubrics for alternative versions are reliable (Rivera & Vincent, 1996). States frequently permit supportive accommodations in order to encourage the participation of English language learners in content assessments in English. Some of the accommodations are designed to reduce the English language demand of assessments for these students by simplifying directions, allowing the use of dictionaries, and reading questions aloud in English. Other accommodations permitted include separate testing sessions, flexible scheduling, allowing extra time, and small group administration (August & Lara, 1996).

Although survey data provided by states indicates the range of accommodations permitted, it is harder to determine which accommodations

BEST COPY AVAILABLE

are actually used. Therefore, states need to collect data documenting how various accommodations are used and how effective they are in promoting the participation of English language learner students in statewide assessments. Rivera and Vincent (1996) caution that accommodations do not work equally well for all English language learners because of wide variations in English language proficiency. While accommodations may make a positive difference for English language learners who already are fairly proficient in English, for those who have very little proficiency in English, they may not make enough of a difference to enable students to perform at high levels.

An issue for states is whether the results of tests taken with accommodations can be compared to the results of tests taken without accommodations. The issue of consistency or comparability across tests will not be resolved easily from a technical standpoint. Because of this, it has been suggested that the assessment results of students who take an assessment without accommodations should be separated from those of students who take the assessment with accommodations. The use of alternative assessment strategies and accommodations requires research to determine their comparability to the assessments used to measure the progress of fluent English speakers. States can learn from empirical studies, conducted by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), that examined the inclusion of LEP students in the National Assessment of Educational Progress (NAEP). Two new research efforts by CRESST researchers focus specifically on the validity of accommodations and modifications in assessments. The primary goal of this research is to produce a continuum of accommodations and modifications that will be appropriate and feasible for use in NAEP. The findings of these CRESST studies should have important implications for the large-scale use of assessment accommodations and modifications.

BEST COPY AVAILABLE

Ensure that translated assessments are equivalent to the English version of the assessment.

National statistics show that approximately 73 percent of students categorized as having limited English proficiency come from Spanish language backgrounds (August & McArthur, 1996). Some states with large and stable populations of these students are developing Spanish versions of content area assessments that can be offered as an assessment option. However, while the limitations of English-only assessment are becoming increasingly obvious, translating a test from one language to another raises many new issues. Because concepts and terminology do not have perfect equivalents in different languages, translated items may exhibit psychometric properties substantially different from those of the original English items. Thus, a translated test may not effectively test the same underlying concepts and competencies (Cabello, 1984; Farr & Trumbll, 1997; Olmeda, 1981). Also, because some languages, such as Spanish, have many dialects, it can be very difficult to translate material in a way that will be similarly understood by most speakers of the language (Estrin, 1993). The difficulty presented by translation was noted in the "Standards for Educational and Psychological Testing":

> Psychometric properties cannot be assumed to be comparable across languages or dialects. Many words have different frequency rates or difficulty levels in different languages or dialects. Therefore, words in two languages that appear to be close in meaning may differ radically in other ways important for the test use intended. Additionally, test content may be inappropriate in a translated version. (AERA, 1985, p. 73)

Furthermore, problems occur in developing effective native-language assessments because many English language learners have limited literacy and language skills in their primary languages and therefore need to use both the native-language version and English language version of the test. Developing and validating equivalent "bilingual" versions of a test (two versions side-by-side) is very difficult. For example, research results from

BEST COPY AVAILABLE

the 1995 NAEP field test of mathematics, which tested items in Spanish-only or in side-by-side Spanish-English formats, illustrate the challenge of using native language or bilingual versions of assessments (Anderson, Jenkins, & Miller, 1996). "This research found substantial psychometric discrepancies in students' performance on the same test items across both languages, leading to the conclusion that the Spanish and English versions of many test items were not measuring the same underlying mathematical knowledge" (August & Hakuta, 1997, p. 122).

Because direct translation may actually introduce more language bias, the most highly recommended procedure in test translation is back translation. In this procedure, the test that has been translated into the second language is translated back into English language. The two English versions are compared, and items showing apparent discrepancies in vocabulary, phrasing, or meaning are modified further in the translated version. When this process is completed, the newly-revised version goes through another back translation. At least three back translations, each conducted by a different translator, are generally recommended in order to prepare a translated assessment that does not introduce discrepancies in meaning inadvertently (Lam, 1991).

Disaggregate assessment data to monitor the achievement of English language learners.

Statewide assessment results should be disaggregated to determine how English language learners are performing as a group. The reporting of disaggregated data at state and district levels will allow for an understanding of the academic development and achievement trends of English language learners and enable local educators to make more meaningful judgements about the effectiveness of instructional programs (LaCelle-Peterson & Rivera, 1994). In addition, data collected in state accountability assessments should include background information on English language learners such as their primary language and the length

of time they have received content instruction in English and instruction in English as a second language.

Although the findings from national studies address issues faced by state policymakers, "[M]any challenges still exist that may stand in the way of best measurement practice and the proper implementation of assessment methodologies that are technically sound" (Olson & Goldstein, 1997). Therefore, further research that involves state and local policymakers, educational leaders, and key constituents is needed. "Different types of large-scale assessments are in use in many different localities, some with very different approaches and purposes than NAEP. Because there are limits to the answers that can be found from the ongoing collection of studies, more research is needed at the national, state, and local levels" (Olson & Goldstein, 1997, p. 76).

What Policies and Practices Should School Administrators and Teachers Follow When Assessing the Academic Performance of English Language Learners?

LaCelle-Peterson & Rivera point out that the question "how should we assess English language learners" has no definitive answer, adding that the best assessment policies will result from "the establishment of processes for experimenting and reviewing assessment strategies in light of the changing English language learner population entering the schools" (LaCelle-Peterson & Rivera, 1994, p. 70). Research and effective practice suggest the following principles and approaches as guidelines for school district administrators and teachers in making decisions about assessing the academic performance of English language learners.

Establish assessment policies before selecting or developing measures. In their comprehensive volume on assessing diverse learners, Farr and Trumbull (1997) argue that school administrators and teachers should

establish policies for their assessment programs before they begin to choose, design, refine, or develop a set of measures that will constitute these programs.

> They should reflect on and discuss the purposes of assessment and the questions they most want to answer about their students. They must think about the potential of any test or assessment measure to interfere with learning or to harm students; they must discuss how much intrusiveness they want to permit and how to integrate assessment with instruction. They must decide what types of assessment are appropriate for which students and what administration procedures they need to follow or modify to accommodate diverse learners. (Farr & Trumbull, 1997, p. 201)

Today, assessment policies are often tied to learning standards and linked to instruction. Schools with linguistically and culturally diverse student populations therefore need assessment policies that draw on teacher commitment to standards, understanding of the purposes of assessment, and knowledge about how culture and language affect learning. Building effective policies requires a consideration of the range of measures needed to assess diverse learners, the many factors that might affect the performance of particular student populations, and how these factors can be addressed through accommodations.

Provide English language learners with instruction that will enable them to develop higher order proficiencies.

English language learners must have adequate opportunities to develop proficiencies based on high learning standards. This means ensuring they have been exposed to challenging learning situations and the full range of desired educational outcomes. They should be thoroughly grounded in what is expected of them, provided opportunities to learn the content being assessed, and taught in ways that will enable them to respond to complex and cognitively demanding tasks (Navarrete & Gutsky, 1996; Navarette, 1994). Most importantly, they must have equitable access to the educational resources and high-quality teachers that will support them

in learning and achieving at high levels (Baker & O'Neil, 1995; LaCelle-Peterson & Rivera, 1994).

Use assessments that are appropriate for English language learners.

Teachers should begin planning for the assessment of English language learners with two questions in mind: 1) What do I need to know about individual children's literacy and language development in order to plan their instruction and assess their performance? 2) What activities and tasks can I use to determine this information? (Farr & Trumbull, 1997). The criteria in Figure 3 (see page 101) should be considered in determining the appropriateness of assessments for English language learners.

Use authentic assessments that draw on English language learners' real-life situations.

Authentic assessments connected to real-life situations will help English language learners understand and apply essential concepts, knowledge, and skills (O'Malley & Pierce, 1996). When developing assessments for culturally diverse student populations, educators should consider how students' life experiences will affect their responses to assessments. English language learners have difficulty learning from and responding to assessment tasks that lack a meaningful context. It is far more likely that these students will develop an understanding of academic concepts if assessment tasks connect to their frame of reference and their personal experiences (Farr & Trumbull, 1997; Koelsch et al., 1995). As Baker notes, in developing more equitable assessments "schools must find ways to deal with children from cultures, languages, and expectations that mainstream America barely understands, if at all" (Baker, 1994, p. 199). Thus, administrators and teachers who wish to develop authentic and meaningful assessments for English language learners should draw on such students' home and community experiences.

FIGURE 3

Criteria for Determining the Appropriateness of Assessments for English Language Learners

1. The extent of ELLs' experiences with the concepts, knowledge, skills, and applications represented in the assessment.
2. The language demands of tasks, particularly for tasks emphasizing higher-order thinking skills.
3. Whether assessment tasks include concepts, vocabulary, and activities that would not be familiar to students from a particular culture.
4. Whether the standards for performance are known and familiar to the ELLs who are being assessed, and whether they understand the processes and products of learning that are valued in the assessment.
5. The prior knowledge and understanding required of them in order to make sense of assessment tasks.
6. Whether they will be able to connect their cultural backgrounds and their experiences to what is expected in an assessment task.
7. Whether the range of assessment tasks are multidimensional in ways that accommodate different culturally-based cognitive styles and modes of representing understanding.
8. Whether the ELLs being assessed have had experience with the format of the assessment.
9. The types of accommodations that will be necessary to give them the same opportunity as other students to demonstrate what they know and can do.

(Baker & O'Neil, 1995; Estrin & Nelson-Barber, 1995; Farr & Trumbull, 1997; García & Pearson,1994; LaCelle-Peterson & Rivera, 1994; Neill, 1995)

Use multiple assessment strategies so English language learners have a wide range of options when showing what they know and can do.

Using a variety of assessments is especially critical for English language learners who need to demonstrate their progress in both language and academic areas over time. They must be given multiple opportunities to show *how* they learn, and to demonstrate *what* they have learned in ways that are comfortable for them and reflect their communication capabilities. This approach is widely supported in the literature. Wiggins (1989) highlighted the importance of variety and flexibility in assessment, emphasizing that assessments should accommodate students' learning styles, aptitudes, and interests. Farr and Trumbull (1997) also emphasized

that using varied approaches that accommodate different learning styles will yield more meaningful results. Latitude should also be given in the time allowed to complete assessment tasks, allowing English language learners time to experiment, draft, reflect, and revise their work (LaCelle-Peterson, & Rivera, 1994; Navarette & Gutsky, 1996). The use of multiple assessments over time will yield a more valid profile of what English language learners have learned to emerge. Allowing them to demonstrate their competence in a variety of ways will yield a deeper understanding of their approach to learning situations, their knowledge of content, and their thinking skills. The use of varied strategies will be important for teachers as well because it will enhance their ability to determine English language learners' progress across a wider range of learning areas, and enrich their awareness of cultural differences in how their students approach learning (Farr & Trumbull, 1997).

Establish scoring criteria for performance assessments that are appropriate for English language learners.

Because performance assessments require teachers to apply clearly defined criteria when determining the level of proficiency a student has demonstrated in responding to a task, special attention must be given to whether scoring criteria provide the basis for a fair evaluation of the responses of English language learners. If scoring rubrics used to assess these students are to be fair, they must be developed by district and school staff who are knowledegable about the linguistic and cultural characteristics of these students and who understand how language and culture influence learning. Performance criteria used to assess English language learners are likely to be unreliable if they are developed by staff who hold views of quality performance that conflict with the understanding of specialists who are most knowledgeable about teaching linguistically and culturally diverse children (Baker & O'Neill, 1995). It is especially important that the role of language be explicitly considered when

BEST COPY AVAILABLE

developing scoring criteria so that English language learners are not penalized inappropriately for lacking Engish language skills. Content-area performance assessments should be scored based on the knowledge, skills, and abilities being assessed, not on the quality of the language in which the response is expressed (LaCelle-Peterson & Rivera, 1994).

The background and expertise of scorers can affect an assessment's fairness and validity, and staff must have adequate expertise and training in order to score the performance responses of English language learners fairly. Furthermore, to understand an English language learner's assessment results, an evaluator must be familiar with the student's cultural and linguistic background as well as the extent to which the student is acculturated (Hernandez, 1994). Teaching staff and specialists can benefit from working as a team in scoring the work of English language learners because by doing so they can deepen their understanding of the relationships between performance standards and effective instructional strategies for these students.

Provide professional development for teachers.

New forms of standards-based assessment require that teachers develop new skills and see their role in new ways. Teachers must be able to build instruction around performance tasks, organize learning around holistic concepts, guide student inquiry, provide a variety of opportunities for students to explore concepts and problem situations over time, use multiple forms of assessment to gather evidence of student proficiencies, and make informed judgments about student progress (Lachat, 1994). For teachers to support the use of alternative assessments with English language learners, they must be proficient not only in subject matter knowledge and current theories of how students learn, but also in knowledge of how language and culture influence student learning and performance. Therefore, professional development for teachers is essential to using

BEST COPY AVAILABLE

alternative assessments with English language learners. "Alternative assessments, with such a premium placed on teacher judgment, make sense only under the assumption that high levels of professional knowledge—about subject matter, language, culture, and assessment—are widely distributed in the profession. Thus, the implications for professional development are very serious" (García & Pearson, 1994, p. 379).

Estrin (1993) noted that teachers need opportunities to learn more about how language and culture affect the classroom and also about particular cultural communities. She suggested that professional development might address such areas as differences in English language learners' communication and cognitive styles, evaluating the language demands of classroom tasks, including all students in classroom discourse, determining students' language proficiencies, and working with different cultural communities.

Implications For Policy and Practice

Increase the participation of English language learners in national, state, and district assessment programs.

English language learners must be included to the fullest extent possible in assessment programs that allow schools, districts, and state education departments to monitor their achievement. Inclusion is essential for determining these students' proficiency in core subject areas, the effectiveness of their instructional programs, and the improvements needed to raise their performance levels. States need to develop common, consistent policies on how to use assessment alternatives and accommodations effectively when testing English language learners who possess varying levels of English proficiency. They also need to ensure the technical quality of translated tests. Guidelines on how to include data on the progress of

BEST COPY AVAILABLE

English language learners in state accountability reports are also needed (August & Lara, 1996; Rivera , Hafner, & LaCelle-Peterson, 1997).

Address the issues raised by including English language learners in state assessments.

Assessment reform that benefits monolingual English students will not automatically benefit English language learners. Therefore, the unique needs of English language learners must be addressed when new statewide assessments are being developed or efforts to raise these students' levels of performance will not succeed. If English language learners are not included in the population sample used for validation, the assessment will not be valid for these students and cannot assess them fairly (LaCelle-Peterson & Rivera, 1994). Technical and measurement issues must be addressed in determining the technical adequacy of large-scale assessments for English language learners. Consideration must be given to whether the assessment provides both a fair opportunity for all students to answer questions across the range of difficulties being tested and whether it provides a reliable and consistent measure of the performance of English language learners. The use of assessment alternatives and accommodations must be examined to determine whether they yield results that are comparable to the assessments used with fluent English speakers.

Give high priority to equity when assessing diverse groups of students.

Poor test performance is a very serious matter for children. Tests can influence children's self-perceptions and others' perceptions of a child's abilities and can lower expectations for achievement (Meisels et al., 1995). In proposing that assessments should accommodate diversity in learning styles, aptitudes, and interests, Wiggins (1989) asked why all students must be tested in the same way and at the same time, and why a student's speed of recall should be so well-rewarded and slow answering so heavily

penalized. This equity perspective highlights the importance of ensuring that English language learners are not penalized because assessments are not fair or appropriate for them or because they are deprived of the time they need to complete an assessment. Equal attention must be given to providing the types of accommodations that will allow English language learners the opportunity to demonstrate what they know and can do (García & Pearson, 1994; LaCelle-Peterson & Rivera, 1994).

Provide English language learners with high-level instruction so they can perform complex and cognitively demanding assessments tasks. Educational policies and practices need to create the conditions that will allow English language learners to achieve high levels of performance. The standards and criteria for performance must be clear to these students, and they must be thoroughly prepared for what is expected of them. They must also be given opportunities to practice and refine their responses to higher-order assessment tasks, and helped to understand the processes of learning that are valued and how they can demonstrate the quality of their work. Clear expectations will help English language learners adjust their performance to the demands of assessment tasks, but more attention must also be given to devising assessment tasks and instructional strategies with diversity in mind. As assessment tasks increasingly measure students' higher-order thinking skills rather than their retention of facts and fragments of information, much more consideration must be given to the influence language and culture have on how students solve problems, make inferences, question assumptions, communicate mathematically, and demonstrate other behaviors associated with higher-level cognitive abilities (Farr & Trumbull, 1997; Malcom, 1991).

BEST COPY AVAILABLE

Use assessment and instructional practices that enable English language learners to connect their cultural backgrounds to the academic knowledge valued in schools.

English language learners can demonstrate what they know and can do more effectively when instruction and assessment draw upon their real-life experiences, allowing them to build upon their prior knowledge and choose their own ways of solving problems. To create meaningful learning contexts for English language learners, educators must understand how instruction and assessment can connect to the cultural experiences of these students. In more concrete terms, assessments should provide a range of options for students to express their knowledge and understanding. Students' home and community experiences can be incorporated into instruction, and learning tasks can include topics that are relevant to students from diverse cultural backgrounds (Farr & Trumbull, 1997; García & Pearson, 1994; Neill, 1995; Winfield, 1995).

Use multiple measures to make decisions about the academic progress of English language learners.

When English language learners have an opportunity to show their understanding and competence in a number of ways, it is more likely that they will be able to demonstrate what they know and can do. Therefore, by using multiple assessments, we can increase the likelihood that our judgments about the progress of these students are valid (Navarrete & Gutsky, 1996). Furthermore, by using multidimensional assessments throughout the school year, teachers and school administrators will be able to get a more meaningful profile of English language learners' language development and academic progress. The use of multidimensional assessments over time can reduce the negative consequences that occur for English language learners when decisions about their achievement and potential are based on information from limited measures that may be susceptible to bias (Farr & Trumbull, 1997).

BEST COPY AVAILABLE

Provide teachers with professional development in how to use performance assessments with students from diverse backgrounds.

Standards-based instruction and the use of alternative assessments with English language learners require new roles and new skills for teachers. Professional development will be necessary to prepare teachers both to build instruction and assessment around authentic learning tasks for students with varying levels of English proficiency, and to evaluate the language demands and cultural content of instructional activities. Professional development needs to strengthen teachers' understanding of how language and culture influence student learning and how differences in the communication and cognitive styles of various cultures influence student participation in learning tasks. Teachers must also receive specific guidance on how to provide a variety of opportunities for students to explore concepts and problem situations over time, how to use multiple forms of assessment to gather evidence of student proficiencies, and how to create and apply scoring rubrics that are not culturally biased.

References

American Educational Research Association, American Psychological Association, National Council of Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anderson, N.E., Jenkins, F.F., & Miller, K.E. (1996). NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics. Princeton, NJ: Educational Testing Service, 1996.

August, D., & Hakuta, K. (1993). Federal education programs for limited English proficiency students: A blue print for the second generation. (Report for the Stanford Working Group). Stanford, CA: Stanford University.

August, D., & Hakuta, K., Editors (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.

August, D., & Lara, J. (1996). *Systemic reform and limited English proficient students*. Washington, DC: Council of Chief State School Officers.

August, D., & McArthur, E. (1996). *Proceedings of the conference on inclusion guidelines and accommodations for limited English proficient students in the national assessment of educational progress, December 5-6, 1994*. National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education. Washington, DC.

Baker, E.L. (1994). Issues in policy, assessment, and equity. In *Proceedings of the second national research symposium on limited English proficient (LEP) student issues: Focus on evaluation and measurements, Vol. II*, pp. 1-18. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs. In B. Farr and E. Trumbull (Eds.), *Assessment Alternatives for Diverse Classrooms*. Norwood, MA: Christopher-Gordon Publishers, Inc.

Baker, E., & O'Neil, H. (1995). Diversity, assessment, and equity in educational reform. In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment*. Boston: Kluwer Academic Publishers.

Banks, J., & Banks, C.M. (Eds.). (1993). *Multicultural education: Issues and perspectives* (2nd ed.). Boston: Allyn & Bacon.

Cabello, B. (1984). Cultural interface in reading comprehension: An alternative explanation. *Bilingual Review, 2*, 12-20.

Calfee, R., & Perfumo, P. (1993). Student portfolios: Opportunities for revolution in assessment. *Journal of Reading, 36*, 532-537.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*, 5-30.

Downing, J. (1984). A source of cognitive confusion for beginning readers: Learning in a second language. *The Reading Teacher, 37*, 366-370.

Estrin, E. (1993). *Alternative assessment: Issues in language, culture, and equity*. San Francisco: Far West Laboratory.

BEST COPY AVAILABLE

Estrin, E.T., & Nelson-Barber, S. (1995). *Issues in cross-cultural assessment: American Indian and Alaska Native students*. (Far West Laboratory Knowledge Brief No. 12). San Francisco: Far West Laboratory.

Farr, B., & Trumbull, E. (1997). *Assessment alternatives for diverse classrooms*. Norwood, MA: Christopher-Gordon.

García, G.E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic students. *Reading Research Quarterly, 26*, 371-392.

García, G.E. (1992). *The literacy assessment of second-language learners* (Tech. Rep. No. 559). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

García, G., & Pearson, P. (1991). The role of assessment in a diverse society. In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 253-278). New York: Teachers College Press.

García, G., & Pearson, P. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of Research in Education 20*. Washington, DC: American Educational Research Association.

Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B.R. Gifford & M.C. O'Conner (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37-76). Boston: Kluwer Academic Publishers.

Glaser, R. (1987). The integration of instruction and testing: Implications from the study of human cognition. In D.C. Berliner and B.V. Rosenshine (Eds.), *Talks to teachers*, (pp. 329-241). New York: Random House.

Gordon, E.W. (1992). *Implications of diversity in human characteristics for authentic assessment*. Berkeley, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hafner, A., & Ulanoff, S. (1994). Validity issues and concerns for assessing English learners: One district's approach. *Educational and Urban Society, 26*, 367-389.

Hernandez, H. (1989). *Multicultural education: A teacher's guide to content and process*. Columbus, OH: Merrill.

Hernandez, R. (1994). Reducing bias in the assessment of culturally and linguistically diverse populations. *Journal of Educational Issues of Language Minority Students, 14*, 269.

Jiménez, R.T. (1992). *Opportunities and obstacles in bilingual reading*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Jiménez, R.T., García G.E., & Pearson P.D. (1991, December). *The strategic reading processes of bilingual Hispanic children who are good readers*. Paper presented at the National Reading Conference, Miami, FL.

Koelsch, N., Estrin, E., & Farr, B. (1995). *Guide to developing equitable performance assessments*. San Francisco: WestEd.

LaCelle-Peterson, M., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review, 64* (1), 55-75.

Lachat, M.A. (1994). *High standards for all students: Opportunities and challenges.* South Hampton, NH: Center for Resource Management, Inc.

Lam, T. (1991). Testing of limited English proficient children. In K. E. Green (Ed.), *Educational testing: Issues and applications.* New York: Garland Publishing Company.

Malcom, S.M. (1991). Equity and excellence through authentic science assessment. In G. Kulm and S. Malcolm (Eds.), *Science assessment in the service of reform,* (pp. 313-330). Washington, DC: American Association for the Advancement of Science.

Medina, N., & Neill, D.M. (1990). *Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in America's public schools* (3rd ed.) Cambridge, MA: FairTest.

Meisels, S.J., Dorfman, A., & Steele, D. (1995). Equity and excellence in group-administered and performance-based assessments. In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Navarrete, C.J. (1994). Developing effective education models for limited English proficient students. In J. Lara (Facilitator), *Assessment of limited English proficient students.* Panel discussion conducted at CCSSO's National Conference on Large-Scale Assessment, Albuquerque, NM.

Navarrete, C.J. & Gustke, C. (1996). *A guide to performance assessment for linguistically diverse students.* Albuquerque, NM: Evaluation Assistance Center-West.

Neill, M. (1995). Some prerequisites for the establishment of equitable, inclusive multicultural assessment systems. In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Olmeda, E.L. (1981). Testing linguistic minorities. *American Psychologist, 36,* 1078-1085.

Olson, J.F., & Goldstein A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* National Center for Education Statistics, U.S. Department of Education, Washington, DC.

O'Malley, J. M., & Pierce, L. V. (1996) . *Authentic assessment for English language learners.* Reading, MA: Addison-Wesley Publishing Company.

Rivera, C., Hafner, A., & LaCelle-Peterson, M. (1997). *Statewide assessment programs: Policies and practices for the inclusion of limited English proficient students.* Paper presented at annual meeting of the National Conference on Large-Scale Assessment, June 24, 1996, Phoenix, AZ. Washington, DC: George Washington University Center for Equity and Excellence in Education.

Rivera, C., & Vincent, C. (1996). *High school graduation testing: Policies and practices in the assessment of limited English proficient students.* Paper presented at annual meeting of the National Conference on Large-Scale Assessment, June 24, 1996, Phoenix, AZ. Washington, DC: George Washington University Center for Equity and Excellence in Education.

BEST COPY AVAILABLE

Rueda, R., & García, E. (1992). *A comparative study of teachers' beliefs about reading assessment with Latino language minority students.* Santa Cruz: National Center for Cultural Diversity and Second Language Learning, University of California at Santa Cruz.

Saville-Troike, M. (1991). *Teaching and testing for academic achievement: The role of language development.* National Clearinghouse for Bilingual Education FOCUS: Occasional Papers in Bilingual Education, No. 4. Washington, DC: NCBE.

Shavelson, R.J., Baxter, & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21*, 22-27.

Valdez-Pierce, L., & O'Malley, J.M. (1992). *Performance and portfolio assessment for language minority students.* Washington, DC: National Clearinghouse on Bilingual Education.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46*, 41-47.

Winfield, L. (1995). Performance-based assessments: Contributor or detractor to equity? In M. Nettles & A. Nettles (Eds.), *Equity and excellence in educational testing and assessment.* Boston: Kluwer Academic Publishers.

The Northeast and Islands Regional Educational Laboratory
A Program of The Education Alliance at Brown University

Adeline Becker
Executive Director, The Education Alliance

Phil Zarlengo
Executive Director, the LAB at Brown University

LAB Board of Governors

Vincent Ferrandino
Chair

Marjorie Medd
Vice Chair

Board Members

J. Duke Albanese
Barbara Bailey
Pamela Berry
James Connelly
Rudolph Crew
Paul Crowley
David Driscoll
Katharine Eneguess
Victor Fajardo
Charlotte K. Frank
Marc Hull
Edward McElroy
Peter McWalters
Richard Mills
Thong Phamduy
Daria Plummer
Arthur Robbins
Olga Lucia Sallaway
Theodore Sergi
David Sherman
Ruby Simmonds
Jeanette Smith
Jill Tarule
Elizabeth Twomey
David Wolk

BEST COPY AVAILABLE